ECONOMIC ANALYSIS OF THE LINK BETWEEN ADVANCES IN BIOPHARMACEUTICAL MANUFACTURING AND HEALTHCARE DELIVERY COSTS

Ebenezer Kwabena Tetteh

Programme of Study:
Research Dept: Epidemiology and Public Health

Department:
Institute of Epidemiology and Health Care

Acknowledgments and disclaimer

This research benefited from a grant disbursed under the UCL Grand Challenges on Human Wellbeing. Additional funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies is gratefully acknowledged. Financial support from the consortium of industrial and governmental users is also acknowledged. Thanks to my supervisors Professor Steve Morris (UCL Department of Applied Health Research) and Professor Suzanne Farid (UCL Department of Biochemical Engineering) – and Professor Nigel Titchener-Hooker (UCL Department of Biochemical Engineering) – for their valuable inputs.

The statements, views and conclusions expressed in this thesis are mine and do not necessarily reflect those of my sponsors or affiliated institutions (in the past or present). All errors are of my own making.

UCL Doctorate in Health Economics

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Name: **Ebenezer Kwabena Tetteh**

Date: 31/07/2016

Abstract

Concerns about increasing expenditures on new biologic medicines has speared interests in finding solutions to contain healthcare delivery costs whilst, at least, maintaining population health indicators at the same levels. Often overlooked are investments in pharmaceutical manufacturing research, which perhaps is because manufacturing costs are thought to only make a small fraction of drug prices, and hence any impact on healthcare delivery costs is small enough to be ignored. This thesis provides evidence, to the contrary, that the impact on healthcare delivery costs cannot be trivialized.

Starting with a systematic review of literature, this thesis finds that there are inconsistencies in what items are included or excluded from estimates of drug administration costs. On the back of this finding, this thesis evaluated drug administration costs for a sample of biologic drugs to identify manufacturing choices that explain most the observed variation in administration costs. It presents an algorithm pharmaceutical manufacturers could use to predict the expected administration-cost savings from reformulating biologic drug candidates still in development and/or existing medicines already on the market. This thesis goes further to argue that a primary focus on the monetary costs of drug administration ignores intangible benefits from satisfying end-user preferences for different modes of drug administration. The monetary value of these intangible benefits could be in the same order of magnitude as savings on the direct monetary costs of drug administration.

In addition, this thesis shows that using cost-reducing process innovations in drug manufacturing, as opposed to cost-increasing ones, is associated with a higher number of products available for clinical use and lower prices over time; whilst the likelihood that a drug product is no longer available is seemingly unrelated to the process innovation deployed in manufacturing. The evidence put together indicate significant societal benefits from developing and implementing innovations in pharmaceutical manufacturing.

Research outputs

Two published papers:

1. "Evaluating the administration costs of biologic drugs: Development of a cost algorithm", *Health Economics Review* 4: 26, 2014
2. "Systematic review of drug administration costs and implications for biopharmaceutical manufacturing", *Applied Health Economics and Health Policy* 11: 445-456, 2013

Two conference presentations:

1. "Pharmaceutical manufacturing, pricing and market conduct" presented at the 10th International Health Economics (iHEA) conference, Dublin, Ireland. 14th – 16th July 2014.
2. "Development of an administration-cost algorithm", presented at the 10th iHEA conference, Dublin, Ireland. 14th – 16th July 2014

Contents

INTRODUCTION

1 Context of Work

1.1 Background

Biologic drugs and related macromolecular therapies (for example, therapeutic proteins, vaccines, monoclonal antibodies, allergy immunotherapies, blood components and fusion proteins as well as gene and stem-cell therapies that are still under research) are a new group of complex products that are derived from living organisms or organic substances and costly to manufacture relative to conventional small-molecule chemically-synthesized medicines. It is often thought that the underlying complex process of manufacturing these medicines, using living cells that vary naturally, is the reason why biologic drugs are pricey, costing up to £70,000 per patient per year in the United Kingdom(UK)[1]. Indeed, high (fixed) costs of manufacturing is one of the reasons why researchers expect less market entry for generic biologics and higher prices relative to chemically-synthesized medicines[2].

It follows that guaranteeing efficient supply and consumption of new biologic medicines will require, among other factors, significant improvements in manufacturing (process innovations) if the UK pharmaceutical industry and National Health Service (NHS) are to realize the full social benefits of their investments in pharmaceutical research and development (R&D), be it process or product innovations. This raises a number of interesting questions: what is the nature of the relationship between manufacturing, the "high" prices and availability of biologic drugs? If there are links between pharmaceutical manufacturing efficiency and healthcare delivery costs, why hasn't previous research considered this issue? One of the reasons why manufacturing has received disproportionately less attention in research on the pharmaceutical industry can be traced to economic theory. This posits that once new medicines become available, sunk or fixed R&D and production costs should have no influence on pharmaceutical prices. Prices of medicines are driven mostly by demand conditions *and* the incremental or marginal costs of production. Since the incremental or marginal costs of production (roughly approximated as cost-of-goods) are often considered a small fraction of total price, most research or analysis of drug pricing ignore manufacturing process improvements[3].

The gaps in research (which this thesis aims to fill) is evidenced by the fact that I found only one study[4] that conducts a detailed economic analysis of the links between the efficiency of pharmaceutical manufacturing, drug prices and public health. This study showed that there are societal benefits from lower manufacturing costs if this translates into

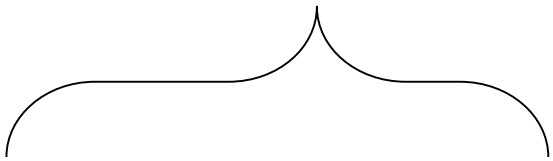
lower prices for payers/ consumers and better access to beneficial medicines in the short-run. Even if lower manufacturing costs are not translated into lower drug prices, there will still be significant societal benefits in the long-run as part of the increase in net sales revenues can be used to finance future R&D in more efficacious new medicines. Whilst this argument is convincing, it is only valid in a healthcare payer environment where, *ex post* of market launch, demand for new medicines are not linked to formal, scientific measurements of cost-effectiveness or value-for-money. It assumes products are demanded once they meet regulatory requirements for safety and efficacy.

In contrast to the situation where healthcare payers are only concerned about the net health benefits offered by medicines they purchase (that is, efficacy minus safety risks), value-based pricing proposed within the UK NHS (sometimes called benefit-based pricing) and presumably intended to be administered by the National Institute for Health and Care Excellence (NICE), aims to ensure prices and demand for new medicines reflect the clinical and economic value to patients and the broader NHS[5],[6]. In such an environment, an investigation of manufacturer's choice of which production process to use and how this affects total healthcare delivery costs; and consequently estimates of cost-effectiveness will be useful for the *ex ante* determination of the manufacturability of biological drug candidates as well as determination of the societal benefits from (public or private) investments in pharmaceutical manufacturing research.

The current approach to investigate these issues is to conduct health economic evaluations, or broadly speaking health technology assessments (HTA). This typically involves scientific measurements of the incremental total healthcare delivery costs and health benefits associated with new medicines relative to existing treatment alternatives.
The diagram below shows the types of evaluations that researchers and analysts usually perform. NICE in most occasions undertakes cost-effectiveness or cost-utility analyses on behalf of the NHS in England and Wales. From this diagram, one is able to identify the manufacturing-related determinants of the costs and cost-effectiveness of health technologies. The variables of interest, and which are amendable to improvements in pharmaceutical manufacturing, are drug acquisition costs and drug administration costs.

The contribution of this thesis to existing literature is therefore to provide empirical evidence and new insights on the societal benefits of innovations in pharmaceutical manufacturing, focusing primarily on healthcare delivery costs.

Health economic evaluation



Resource Cost

C_1 = Direct costs
$C_1 = C_j^0(P_j) + C_j^1(R_j) + \Sigma C_j^2(E_j)$
$C_j^0(P_j)$ = acquisition costs for product j
$C_j^1(R_j)$ = administration costs
$\Sigma C_j^2(E_j)$ = future healthcare-related costs
C_2 = Indirect costs* (productivity losses)
C_3 = Intangible costs*

Health Benefits (expected)

E= Benefits in natural units
U = Benefits in utility units†
B = Benefits in monetary terms (£, $)
B_1 = direct benefits
B_2 = indirect benefits (productivity benefits)*
B_3 = intangible benefits*

Types of economic evaluation

Cost Analysis:	C_1; C_1+C_2
Cost-effectiveness Analysis (CEA):	$(C_1+C_2)/E$; $(C_1 - B_1)/E$; $(C_1+C_2 - B_1 - B_2)/E$
Cost-utility Analysis (CUA):	$(C_1+C_2)/U$; $(C_1 - B_1)/U$; $(C_1+C_2 - B_1 - B_2)/U$
Cost-benefit Analysis (CBA):	$B_1+B_2 - C_1 - C_2$; $(B_1+B_2)/(C_1+C_2)$

Source: Adapted from Mills and Drummond[7] and Drummond[8].
Notes: P*j* refers to price of product *j*, R*j* is its dosage regimen and formulation, and E*j* refers to its clinical effectiveness. *Included or excluded depending on whether a payer or societal perspective is taken. †The measure of interest could be quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs).

In this thesis, I describe changes to how drugs are manufactured as "process innovations" or "improvements in pharmaceutical manufacturing". These changes could lead to an increase or reduction in manufacturing costs (measured as cost-of-goods); and/or changes in healthcare delivery costs. I define healthcare delivery costs as the sum of drug acquisition costs and administration costs. The latter covers indirect/ intangible costs related to preferences for different modes of administering drugs. Acquisition costs here refer to the product of drug prices and demand volumes. I ignore future, post-treatment healthcare-related costs from my analyses as I do not expect advances in pharmaceutical manufacturing to have any significant or generalizable effects on future healthcare costs. The point here is: even if manufacturing improvements or innovations amplify and/or reduce the uncertainty around health benefits derived from a given drug (and hence change the stream of future healthcare-related costs), accounting for these effects will require some clinical-trial or observational evidence that is most likely to be product-specific.

This seemingly simple definition of healthcare delivery costs (as made up of drug acquisition and administration costs), however, provides useful insights on the potential sources of the societal benefits from improvements in pharmaceutical manufacturing. Manufacturers' ability to predict and control the healthcare delivery costs and cost-effectiveness of new medicines (especially for drug candidates that are still in development) may well be the crucial determinant of their ability to recover the costs of sunk R&D investments whilst making adequate competitive profit to maintain their business interest in drug research and development. Improvements in pharmaceutical manufacturing may well be the difference between patients having access or no access to treatments that offer positive net health benefits but are not considered recommended cost-effective options.

1.2 Aims and objectives

The aims of this thesis are to establish and evaluate the links between pharmaceutical manufacturing costs (or more generally speaking, improvements in pharmaceutical production) and healthcare delivery costs, focusing on new and emerging stream of biologic medicines and related macromolecular therapies. Although the focus is on biologic medicines (and emergent macromolecular therapies), it aims to provide outputs that are also of relevance to small-molecule chemically-synthesized medicines.

Given the simplified definition of healthcare delivery costs above, and prior expectations that improvements in pharmaceutical manufacturing may influence healthcare delivery costs via (1) drug acquisition costs and (2) drug administration costs; my objectives are as follows:

1. To investigate and develop new insights on the relationship between pharmaceutical manufacturing (costs) and drug pricing;
2. To conduct a systematic literature review of administration costs for biologic drugs and the implications for biopharmaceutical manufacturing;
3. To develop a regression-based algorithm for the evaluation of drug administration costs; and
4. To conduct a discrete-choice modelling of preferences for different modes of drug administration.

The research objectives listed above are organized under two broad themes: "manufacturing and drug administration costs" and "manufacturing and drug acquisition

costs". The chapters and contents of this thesis are structured to follow the themes and objectives listed above. They should, however, not be considered as separate parts. A brief description of my research plan including the theories, tools and techniques used are given below.

1.2.1 Manufacturing and drug administration costs

Systematic literature review of drug administration costs

The first research stream under "Manufacturing and Drug Administration Costs" is to conduct a systematic review of published, peer-reviewed literature that report estimates of drug administration costs for biologics within the UK NHS. The systematic review is intended to shed some light on how different components of drug administration costs have been quantified and to examine how differences in dosage forms and regimens can influence administration costs. I first developed a framework of drug administration costs; and used this framework to evaluate administration costs for biologics. From the outputs of the literature review, I discussed the implications of my findings for research in biopharmaceutical manufacturing, focussing on how healthcare delivery costs and associated drug formulation issues may affect or influence manufacturing and R&D decisions.

Development of algorithm for evaluating drug administration costs

The second research stream under the theme "Manufacturing and Drug Administration Costs" is a de novo analysis of administration costs for a sample of biologics with UK marketing authorization. The analysis makes use of the framework on drug administration costs above to compute, first, deterministic estimates of drug administration costs and then to perform some Monte Carlo simulations to reflect more closely variation in administration costs in routine clinical practice. Using the simulated outputs, I developed a regression-based cost algorithm that can be used to evaluate: (1) how the administration costs for a given biologic medicine varies across acute and chronic clinical indications; (2) variation in administration costs by route of administration; and (3) differences in drug administration costs for different dosing regimens of a given biologic. This algorithm can be used, within reasonable margins of error, to predict the administration costs of biologic drug candidates that are under development.

Preferences for different modes of drug administration

The third research stream under "Manufacturing and Drug Administration Costs" investigates the hedonic, non-monetary attributes of different modes of drug administration that is preferred by end-users, i.e., patients (otherwise healthy people) or healthcare professionals acting on behalf of patients. It focusses on preferences for modes of drug administration in terms of convenience, continuity of treatment, less disruption of daily activities, avoidance of needles etc. And makes use of a discrete choice experiment (DCE) which is one of a number of methodologies used for the valuation (in monetary terms) of goods and services for which no market exists; or that observed demand curves do not provide a valid measure of benefits (in excess of costs) due to market imperfections and the like.

Following the arguments and recommendations made by practitioners in the field[9],[10],[11], the DCE consisted of the following general steps:

1. Conceptualizing the choice process (within the tenets of random utility theory),
2. Identifying the characteristics (attributes) of different modes of drug administration,
3. Assigning levels to the characteristics (attributes) identified,
4. Developing choice scenarios for the DCE questionnaire,
5. Questionnaire design and piloting,
6. Collecting choice or preference data via responses to the questionnaire, and
7. Econometric analysis of the choice data generated.

I evaluated the choice preferences of (1) people from the UK general public and (2) healthcare professionals from the US – and these were collected from a web-based survey. The inclusion of US healthcare professionals rather than those in the UK simply reflects the resource and time constraints for this research. I considered that having some idea of the preferences of agents acting on behalf of patients, in contrast to evaluating just the choices of patients, will help provide a better picture of the desirable attributes of different modes of drug administration.

1.2.2 Manufacturing and drug acquisition costs

Pharmaceutical manufacturing, market conduct and pricing

This work investigates possible relationships between manufacturing (costs) and pharmaceutical pricing and market conduct. The main objective is to generate some empirical evidence in support of or against the existence of significant societal benefits from advances in biopharmaceutical manufacturing. The existence of these benefits could then be used to

determine whether there is a case for private or public investments in pharmaceutical manufacturing research. I first conducted a selective non-systematic review of literature on the benefits of pharmaceutical manufacturing efficiency; and then used elements of market theory and theories of the firm to describe a framework that highlights the welfare implications of using process innovations that reduce or increase production costs, relative to previous manufacturing processes. This is followed by an empirical econometric analysis to establish the relationship between improvements or innovations in manufacturing and drug prices, the number of products available for clinical use and the hazard of product exit. The empirical analyses (conducted using panel and survival regression methods) were based on NHS data on a sample of biologic medicines over a period of time covering on-patent and off-patent periods. From these analyses, I draw some arguments and conclusions as to whether the manufacturing process innovations embodied by biologic medicines have a bearing on drug pricing, number of products available and how long these products remain available on the NHS market.

1.3 Thesis layout

The rest of this thesis is organized into two broad parts: “Manufacturing and drug acquisition costs” and “Manufacturing and drug administration costs”. The first part consists of only one chapter investigating the links between pharmaceutical manufacturing, market conduct and pricing. The second part is made up of three chapters, describing (1) a systematic review of literature reporting on drug administration costs; (2) a de novo analysis of drug administration costs; and (3) a discrete-choice evaluation of preferences for different modes of drug administration. References and appendices will be found at the end of the thesis.

MANUFACTURING AND DRUG ACQUISITION COSTS

2 Pharmaceutical Manufacturing, Market Conduct and Pricing: Lessons for Future Biologics

2.1 Introduction

The success of biologic drugs[1] in the treatment of indications such as cancer and autoimmune diseases has led to them becoming the fastest growing class of new medicines in the pharmaceutical sector: 40% of drug candidates in R&D are biologics with a worldwide market forecast of circa $145 billion, growing at more than 15% annually [12]. However, from a healthcare payer perspective, biologics are among the most expensive drugs available: for example, the annual acquisition cost per patient can exceed £70,000 for monoclonal antibody treatments for cancer [13]. This is a major financial burden on attempts to ensure wider access to the health benefits provided by these new drugs. This burden has led the UK NHS to use increasingly complex coverage and reimbursement criteria to justify funding for these newer but more expensive drugs. In England and Wales, for instance, NICE uses clinical- and cost-effectiveness criteria to advise the NHS on decisions regarding which drugs to adopt. Cohen et al. [14], for example, report that 40% of health technologies including medicines assessed by NICE are conditionally recommended for restricted use and < 10% are not recommended at all due to concerns about clinical and cost effectiveness. Why are biologic drugs so expensive and what can be done to make them affordable?

The high acquisition cost of biologic drugs can be justified (1) as rewards for the incremental benefits they offer, i.e., measured clinical or economic value – and (2) as contribution margins towards the cost of manufacturing, R&D expenses on successful and failed drugs; selling, general and administrative expenses; and a reasonable profit return. Cost of goods (COG), i.e., the cost to manufacture biologics is reported to typically represent 15-40% of the selling price [15;16],[17]. However, the cost of manufacturing biologic drugs need not be or remain 'high'. Older biologically-derived pharmaceuticals such as antibiotics introduced in the 1940s and vaccines were at first hard and expensive-to-manufacture but then experienced significant price reductions over time when they became easier and cheaper-to-make. Of course, price reductions could be due to changes in output (i.e., widened

[1] In this chapter "biologic drugs" are defined as active substances or medicinal products derived from a biological living system (of plant, animal or microbial origin), irrespective of how small or large the molecules are. This definition covers not just recombinant proteins, monoclonal antibodies, blood and immunological products (vaccines, sera, allergens etc.), advanced technologies such as gene and cell therapies – but also semi-synthetic products of biological origins that do not belong to any of the groups listed above.

patient access to medicines) but the question I ask is: are there manufacturing lessons to extract from these older antibiotics and vaccines and possibly extended to the current cohorts of biologic drugs?

This chapter explores how manufacturing-cost reductions over time might contribute to changes in market conduct that provide significant societal benefits in terms of affordable medicine access and security of supply. It focuses on the number of products available (what I call product density), the hazard of product exit and variation in product prices over time. In particular, it tests whether, through improvements in manufacturing, existing and emergent biologic drugs (such as monoclonal antibodies) may go through the transition from being 'hard-to-manufacture' and 'expensive-to-buy' to products that are 'easy-to-make' and 'less-costly-to-buy'. My analyses are based on data from the UK NHS for the following biologically-derived pharmaceuticals: penicillins, hepatitis B vaccines, influenza vaccines, poliomyelitis (hereafter polio) vaccines and the measles-mumps-rubella (MMR) vaccines. I focus on these drug classes because the principles behind the methods used to manufacture the selected sample products are similar to or formed the bases of production processes used to make current and emerging cohorts of biologics. This chapter, however, concentrates on only the short-run impact improvements in pharmaceutical manufacturing may have on market structure and conduct. It does not consider the long-run impact of process innovations on the intensity of product R&D.

This chapter is structured as follows: section 2.2 provides a selective review of existing literature on the value of manufacturing process innovations and describes the theoretical framework underlying my analysis. Section 2.3 describes the data and statistical methods used in this study and section 2.4 presents my results. I discuss in section 2.5 my findings and present my conclusions in section 2.6.

2.2 Value of pharmaceutical process innovations

2.2.1 Related Work

Vernon, Hughen and Trujillo[4] report significant societal benefits of improving pharmaceutical manufacturing efficiency from a short-run or long-run perspective. The stream of benefits depend crucially on how manufacturing efficiency gains alters the structure of and conduct in drug markets. The paper focuses on two boundary scenarios in which

society can benefit from improvements in pharmaceutical manufacturing. In the first scenario *A*, lower manufacturing costs in the short-run translate into lower short-run (profit-maximizing) prices whilst keeping price-cost margins the same. In what is supposed to be the more realistic scenario *B*, efficiency in manufacturing does not lead to a fall or rise in prices but short-run price-cost margins increase. Scenario *B* reflects the notion that there is no guarantee that pharmaceutical companies will pass on manufacturing-cost savings as lower prices for consumers or healthcare payers acting on their behalf – especially when competitive pressures are absent, non-functional or imperfect. That is to say, given prevailing demand conditions, the structure of the supply market is critical to whether changes in COG will result in consumer price changes. The authors, however, argue that while price-cost margins in scenario *B* may rise in the short-run, society still stands to gain in the long-run from increased investments in product innovations used by consumers and their households. They cite Scherer [18], who reported a cyclical relationship between gross price-cost margins and the intensity of pharmaceutical R&D in product innovation. Higher gross price-cost margins in the short-run are associated with higher R&D expenditures and vice versa; a correlation of +0.92.

This is consistent with other studies. Basu et al. [19], using published data in the annual reports of twenty-two US pharmaceutical companies over the period 1994 to 2005, report that COG for brand-name ethical companies exhibited a significant declining trend whilst R&D expenses exhibited an increasing trend over the same period. Correlation between COG and R&D expenses is reported to be –0.93 for brand-name companies. Ethical pharmaceutical firms do appear to reserve, at least, some of the savings on COG for product R&D rather than appropriate all cost-savings as pure profits. The same argument, however, cannot be extended to generic firms as the sort of product R&D they engage in is developing new formulations for existing drugs. For generic companies, the study found a correlation between COG and operating incomes of –0.92: manufacturing cost savings resulted in purely an increase in profit margins. For biotechnology firms, COG and R&D expenses did not change significantly over the specified time period and no clear trend was observed.

The upshot is: irrespective of whether reductions in COG are reflected in price reductions or not, manufacturing process innovations should lead to improvements in consumer or social welfare via either lower prices in the short-run or in the long-run, supply of new products having significant impacts on mortality (life years saved) and/or morbidity (increments in health-related quality-of-life). Lichtenberg's [20] estimation of the contribution of pharmaceutical R&D to US life expectancy suggests for every $926 invested

in R&D, one life year is saved. Using this statistic and a cost-effectiveness threshold of $100,000 per life year gained to compute a monetary value for health benefits, Vernon et al. [4] report that the societal gains from scenario *B* is approximately 12 times that from scenario *A*. However, for the higher price-cost margins realized from process innovation to represent a welfare gain in the long-run, pharmaceutical firms must invest a proportion of 'profits' in product innovation.

2.2.2 Theoretical Framework

Figure 2.1: Price, output and welfare consequences of process innovations

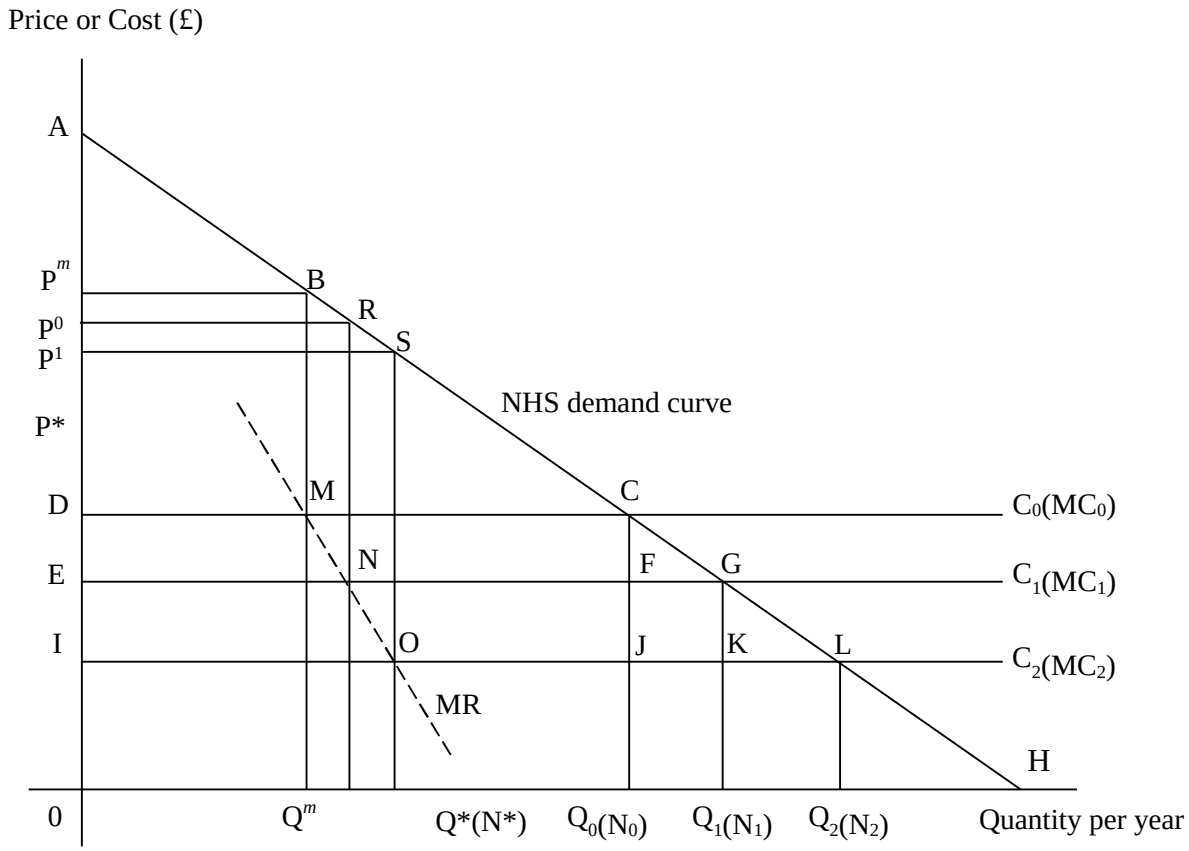


Figure 1 above depicts the private and societal benefits of improvements in pharmaceutical manufacturing. Figure 1 is not drawn to scale and it shows, for simplicity, a linear aggregate NHS demand curve (in a given year) for a given therapeutic class of drug products with similar chemical structure, pharmacologic mode of action or clinical indications for single or multiple illnesses. This is the definition of the 'market' here and the line AH can be considered as a series of short-run NHS demand curves over the lifecycles of the products demanded. Whilst Figure 1 appears similar to the static textbook model of

monopoly, I give it a dynamic interpretation by considering that each therapeutic 'market' goes through three possible stages: (1) a monopolistic class containing a patented drug with no substitutes; (2) a duopoly or oligopolistic class containing patented drugs with streams of similar but differentiated therapeutic substitutes; and (3) a competitive class containing off-patent generic equivalents. See DiMasi and Paquette [21]. I label these stages T_0, T_1 and T_2 respectively.

In each year of the monopolistic stage T_0, a manufacturer supplies a first-in-class drug with a constant average cost of production (C_0) which is equal to marginal costs (MC_0). The manufacturer (within government-imposed limits on drug pricing or total profits earned over a firm's portfolio of products supplied to the NHS) faces a marginal revenue curve MR; charges a profit-maximizing price P^m for supplying Q^m of the patented drug[2]; and captures per year quasi-rents, i.e., revenues in excess of production costs (also referred to as producer surplus) given by the rectangle P^mBMD. Net consumer surplus per year accruing to the NHS and its patient populations is given by the triangle P^mAB. This includes the health benefits derived from the drug supplied.

Let's first focus on the welfare effects of cost-reducing process innovations. If the manufacturer is able to develop and implement a cost-reducing process innovation to lower COG to C_1, it will charge a new profit-maximizing price that is lower than P^m, increase output beyond Q^m and capture per year quasi-rents given by the rectangle P^0RNE, which is greater than P^mBMD. If the manufacturer pursues a constant price-quantity strategy, it will reap incremental quasi-rents given by the rectangle with edges EDM. Net consumer surplus per year then increases by P^mBRP^0. (I assume here that the expected increment in quasi-rents more than offsets the associated fixed costs of developing or reverse-engineering the manufacturing process – otherwise the manufacturer wouldn't consider implementing a process innovation.)

If the manufacturer is able to develop and implement a cost-reducing process that lowers production costs to C_2, it will charge a new lower profit-maximizing price, increase output by a larger amount, and capture a higher volume of quasi-rents per year given by the rectangle bounded by P^1SOI. Introducing a cost-reducing process innovation effectively increases the market size for the manufacturer – once the fixed costs of introducing the process innovations are covered. Net consumer surplus per year is now given by triangle

[2] Depending on current clinical and economic evidence available, such monopolistic first-in-class products may face actual or potential competition from existing products from other therapeutic classes/ markets if these products have licensed or unlicensed uses for treating or managing the same set of illnesses. I assume here that the manufacturer of a first-in-class drug sets its profit-maximizing price taking into account these "out-of-class" alternatives and its residual demand curve.

P^1AS – but observe the fall in profit-maximizing prices is much smaller than the associated fall in COG. The monopolist doesn't offer price cuts but its *profit-maximizing price* falls with successful attempts to lower manufacturing costs. This highlights the problem posed by a 'progressive' monopolist: as the manufacturer gets more successful with process innovation, the deadweight loss per year increases from triangle BMC at C_0 to triangle RNG at C_1 to triangle SOL at C_2, whilst net consumer surplus only increases by the small trapeziums along the line BRS. See Reksulak et al. [22].

Beyond stage T_0, the volume of quasi-rents enjoyed by the first-mover manufacturer under temporal monopoly power will, sooner or later, incite competitive rent-seeking in stages T_1 and T_2 by therapeutic substitutes and in stage T_2 by generic equivalents. The greater the volume of quasi-rents available to be appropriated, the larger the incentive for competitive rent-seeking and the higher the number of competing manufacturers or products attracted to the 'market', especially when it is easy to duplicate the (patented) manufacturing process or reverse-engineer a me-too process. Repeated improvements in manufacturing that lowers production costs will shift the potential competitive price-quantity combination from $P_0(C_0)$-$Q_0(N_0)$ to $P_1(C_1)$-$Q_1(N_1)$ to $P_2(C_2)$-$Q_2(N_2)$ and so forth; where N is the number of products that ensure (near) complete dissipation of all quasi-rents available and its transformation into consumer surplus over periods T_1 and T_2. Holding all else equal, $N_2 > N_1 > N_0$; the inverse price-*N* relationship gets steeper as N increases; the potential net consumer surplus per year increases from triangle ADC to AEG to AIL; and the products that constitute N_2 will embody a (richer) history of cost-reducing process innovations[3].

Note that the increments in net consumer surplus over stages T_1 and T_2 come not only from lower prices (or more cost-effective delivery of health benefits) but having an expanded arsenal of alternative product options to cater for heterogeneous clinical needs. Also, the extent to which prices fall in periods T_1 and T_2 with changes to drug manufacturing will depend on whether competitive pressures are functional and effective. Indeed, a lot more can

[3] The theoretical framework presented captures in a parsimonious, non-mathematical manner events that occur over periods T_0, T_1 and T_2. The behaviour of (progressive) monopolists; the competitive and anticompetitive actions and reactions of manufacturers of equivalent or differentiated products; the effects of market-entry barriers (for instance, patents for process innovations), are all assumed to occur year after year until Q(N) is reached. The time path to the competitive price-quantity outcomes may be characterised by entry of products and firms who *ex ante* expect to earn positive quasi-rents or exit of products and firms earning negative returns *ex post*. If the hazard of product exit exceeds the rate of entry, the market will experience a 'shakeout': a persistent decline in the total number of products until N is reached. One may then observe an inverted (skewed) U-shaped distribution for product density over time; but it is also possible that the repeated stream of process innovations (and demand conditions) will be such that the market will not experience any shakeout until N is reached – in which case one will observe an upward-trending distribution for product density.

be said about the nature of the price-*N* relationship by expressing the inverse demand curve in Figure 1 as:

$$P_{ijc}(t) = D^{-1}\left(Q_{ijc}, P^{ge}\left(N_{gj}\right), P^{ts}(N_C, M_c), \ldots\right) \quad (1)$$

where Q_{ijc} is demand quantity for product *i* in molecule *j* in therapeutic market *c*; N_{gj} is the number of generic products in molecule *j*; N_c is the number of substitute molecules for product *i* in molecule *j*; M_c is the mean number of products for the $N_c - 1$ other substitute molecules on the market; P^{ge} and P^{ts} is the mean price of generic equivalents and therapeutic substitutes respectively. On condition that NHS demand is and remains price-sensitive, $\partial P_{ijc}/\partial N_c < 0,\ \partial P_{ijc}/\partial M_c < 0,\ \partial P_{ijc}/\partial N_{gj} < 0$. If generic equivalents are considered closer alternatives than therapeutic substitutes, then $\left|\partial P_{ijc}/\partial N_{gj}\right| > \left|\partial P_{ijc}/\partial N_c\right|$. If product classes embodying cost-reducing process innovations (abbreviated as "costredinnov") have higher product density, then it follows that $\partial P_{ijc}/\partial \text{costredinnov}_{ijc} < 0$.

One will also expect $\partial P_{ijc}/\partial Q_{ijc} < 0$ as, by definition, being price-sensitive means offering more than equiproportionate increases in demand volumes in return for price discounts. Another factor that might influence the price-*N* relationship is the age of a product. If the incremental clinical value of new products (in terms of better efficacy, safety or both) is substantial, then one will expect a (near) complete shift of demands from older, less effective products to newer, more effective products that often carry a price premium constrained only by the NHS estimation of the value-for-money offered by these products. On the other hand, if the difference in clinical value is 'small', then we will expect price competition to evolve such that over a product's lifecycle, $\partial P_{ijc}/\partial \text{Age}_{ijc} < 0$. Older products will have lower prices as long as they are still demanded and embody a history of repeated efforts to lower COG.

Let's now turn our attention to the welfare effects of cost-increasing process innovations. Initially the manufacturer supplies a first-in-class drug in stage T_0 at a cost of $C_2(MC_2)$; but to address concerns about efficacy and safety, it introduces a cost-increasing process that shifts the average cost of production to $C_1(MC_1)$. Introduction of an even costlier manufacturing process will shift the average cost curve to $C_0(MC_0)$ leading to an increasing schedule of profit-maximizing prices, lower supply quantities and shrinking volume of quasi-rents for the same level of demand. The incentive for competitive rent-seeking in stages T_1 and T_2 is thus dampened such that we will not expect prices to fall to the same extent or the price-*N* relationship to be as steep as in the case of introducing a cost-reducing process

innovation. But if because of concerns about efficacy and safety, a drug product is not used or its level of consumption is restricted below what is prescribed in its marketing authorization, then prior to introducing a cost-increasing process innovation, and at production costs of $C_2(MC_2)$, the quasi-rents a manufacturer can expect to capture in a year will be less than P^1SOI, if not zero. Net consumer surplus per year will also be less than P^1AS, if not zero.

Unless the (fixed) costs of introducing a cost-increasing process innovation exceeds the residual product or firm demand (in which case one would expect product or firm exit), the introduction of a cost-increasing process innovation (that resolves efficacy and/or safety concerns) will actually increase the market value of the manufacturing firm if it captures quasi-rents per year of P^0RNE at production costs of $C_1(MC_1)$. In fact, to the extent that introducing a cost-increasing process innovation leads to a positive and more certain realization of net consumer surplus per year (P^0AR), the manufacturer could charge a premium on its profit-maximizing price – again with the size of the price premium constrained by the NHS willingness-to-pay for health benefits.

Note that even when a cost-increasing process innovation is employed, to displace what was an existing cost-reducing manufacturing process, the period after its introduction typically involves attempts to reduce COG through automation and troubleshooting manufacturing issues, for example. These efforts to reduce COG or supply costs will be complemented by gains from learning-by-doing effects and economies-of-scale. That is, the initial upward shift in manufacturing cost to say $C_0(MC_0)$ from introducing a cost-increasing process innovation will be followed by attempts to bring costs down to $C_1(MC_1)$, $C_2(MC_2)$ and further down if possible. The incentive for competitive rent-seeking (in a product class embodying costly process innovations) should increase as demand uncertainty is resolved; product density should also rise and excluding a (near) complete shift in clinical preferences towards new or improved products, we will also expect $\partial P_{ijc}/\partial Age_{ijc} < 0$. Thus a 'market' that first experiences cost-reducing process innovations and later cost-increasing process innovations will be characterised by a steady decline in the number of products and firms. Conversely, a 'market' that experiences cost-increasing innovations and later cost-reducing process innovations will tend to be characterised by multiplicity of products; and a more downward pressure on prices.

It is worth emphasizing that the diagrammatic exposition above assumes a fixed linear demand curve. The effect of changes to drug manufacturing on the price-*N* relationships, patterns of product exit and net consumer surplus will vary in the case of probabilistic linear demand curves. Consider a shift of the line AH in Figure 2.1 upwards and to right to

represent medicinal demands (per year) for treating a severe disease affecting a larger number of people and for which the expected discounted present value of quasi-rents supports a higher level of R&D in new products. The competitive market outcomes with the use of cost-reducing process innovations will lie beyond $Q_2(N_2)$. Next consider a shift of the line AH downwards and to the left to represent medicinal demands (per year) for treating a rare disease affecting a smaller number of people and for which the expected discounted present value of quasi-rents only supports limited investments in product R&D. In this case, the competitive outcomes with changes to drug manufacturing will lie in the proximity of Q*(N*).

In what follows, I do not attempt to measure (in monetary terms) the size of changes in consumer-welfare described above but seek to empirically evaluate the relationships between the various parameters that determine the size of changes in net consumer surplus over the lifecycle of drug products, i.e., price changes over time, the likelihood (hazard) of product exit, product density (i.e., the number of products available); the number of manufacturing firms, residual demand volumes (per firm) in a given therapeutic market; the number of generic equivalents and therapeutic substitutes; product age; the cost-reducing nature of the process used in drug manufacturing and the availability of imported products that might have different manufacturing-cost profiles.

2.3 Data and methods

2.3.1 Resolving an Estimation Problem

There are two ways of evaluating the welfare implications of changing the manufacturing process embodied by products in a given therapeutic class. The first is what I call the ‘direct approach’ which involves looking at the actual production costs (COG) for a sample of biologically-derived products supplied by a number of pharmaceutical companies over a defined period of time. One can then try to establish a relationship between the trend in production costs with price trends, the number of products or any other metric that is considered appropriate. The problem is: it requires detailed data on the COG for the selected product sample. Typically, there is a lack of data on COG, and if such data were available, it was unlikely to be disclosed by companies because of commercial reasons.

The second is what I refer to as the 'indirect approach'. This (again using a sample of biologically-derived pharmaceuticals) relies on expectations that the density of products will be higher in 'markets' where products embody a history of repeated process innovations to lower COG. And, if this higher product density persist (i.e., the hazard of product exit is low, close to zero or remains constant, in spite of competitive pressures on prices), then the steadiness and magnitude of this trend gives us (greater) confidence of the impact of cost-reducing process innovations. The point is: falling product prices cannot sustain overcrowded therapeutic markets without efforts to lower COG and if manufacturers do not realize some 'profit' margin on production costs, they will exit the market. But if price competition is such that a product class embodying a cost-reducing process innovation experiences a shakeout, i.e., the hazard of exit increases over time, then the *ceteris paribus* effect of process innovation may be confounded, making it more difficult to separate out the effect of process innovations from changes in demand conditions or the intensity of price competition.

This indirect approach (which bears some resemblance to what Stigler [23] calls the 'survivorship technique') tests the following hypotheses:

1. Ceteris paribus, the more a given therapeutic drug class or 'market' is characterized by or embodies cost-reducing (cost-increasing) process innovations, the greater (lower) the number of products that can be sustained in that market;

2. Ceteris paribus, the hazard of exit will be lower (higher) in a 'market' where drug products embodies cost-reducing (cost-increasing) process innovations – even in the presence of competitive pricing; and

3. Ceteris paribus, the higher (lower) the number of products in a 'market' that is characterized by cost-reducing (cost-increasing) process innovation, the more (less) intense competitive pressures will act to lower product prices.

2.3.2 Product Sample and Data Sources

My product sample covers drug classes that I believe are closely related to, or at some point time shared the characteristics of current and emerging cohorts of biologic drugs. The drug classes are: penicillins, hepatitis B vaccines, influenza vaccines, polio vaccines and MMR vaccines. See Appendix 2A.

I considered that each product in the sample embodies either a cost-increasing or cost-reducing process innovation depending on whether their *initial* introduction contributed to an increase or a reduction in COG, relative to the previous process innovation(s). This was based on a review of published and unpublished texts on manufacturing innovations related to the selected product sample. This categorization is summarized in Table 2.1 below – and it was validated in discussions with biochemical engineers and an expert in the development of biologic products. Note that the process innovations described in Table 1 include what Mansfield et al. [24] will classify as "product innovations used by manufacturing firms" to make products (that are used by consumers and their households) as well as "process innovations" to lower the cost of manufacturing.

The downside of categorizing products according to whether they are made with cost-reducing or cost-increasing process innovations is: it says nothing about the magnitude of the reduction or increment in manufacturing costs. I believe, however, that this is a credible way of capturing the cost impact of process innovations in the absence of the desired data on COG for each product over time. From Table 2.1, the penicillins (as a class) embody the richest history of cost-reducing process innovations, with some penicillins made with a number of pre-2000 "old" innovations whilst other penicillins are manufactured with "new" post-2000 technologies. I had no information to indicate which of these innovations were deployed in manufacturing a specific drug product, nor did I have any information to indicate whether *all* these innovations were adopted in a one-step change. Although Table 2.1 provides a timeline of when the process innovations were introduced, I had no information as to whether originators or generic entrants in each therapeutic class took the lead in introducing these process innovations. All that can be inferred from Table 2.1 is whether (a series of) process innovations led to a fall in COG or not. I had no data on the magnitude of the fall or rise in COG of the products in each therapeutic class.

In my analyses, the categorization in Table 2.1 (which is not based on product-specific COG data) is accounted for by a dummy variable labelled COSTREDINNOV to indicate products made with cost-reducing process innovations or not. For example, consider two products *A* and *B* belonging to the same class or therapeutic market. If the newer process by which drug *A* is manufactured is less costly than the older process by which drug *B* is manufactured, then the COSTREDINNOV variable for drug *A* will take the value of one whilst that for drug *B* will take the value of zero. Values for the COSTREDINNOV variable switch from zero to one, one to zero within a *product class* (therapeutic market), not within a drug product. Now because the products that embody the process innovations in Table 2.1

have become generic for some time, it is important to disentangle the effects of "genericisation" from the effects of (a series of) changes to the drug manufacturing process. I do this using regression methods that control for the number of generic equivalents a given product has (see section 2.3.3).

Courtesy of the NHS Health and Social Care Information Centre (HSCIC), I obtained data on products (in the selected therapeutic markets) that were prescribed in the NHS over the period 1991 to 2012. The HSCIC did not have data preceding the year 1991. My dataset therefore only offers truncated snapshots of the economic lifecycle of products in my sample: it contains a heterogeneous mix of incomplete product lifecycles. I collected information on: (1) the net-ingredient-cost per (NIC) item, which I use as a rough proxy for product price; (2) the number of products on the NHS market in each year; (3) the number of manufacturers of each product type; and (4) time of product entry into the NHS market and the time of exit, i.e., when a product is no longer available on the NHS. Data on the NIC per item was specifically chosen to reflect the direct acquisition costs to the NHS: it excludes cost components such as discounts, prescription and dispensing fees and other sources of health-staff income. All product 'prices' have been adjusted to 2011 values using the Hospital and Community Health Services pay and price inflation index. I also collected data on demand volumes (proxied by the number of prescription items) and NHS expenditures (i.e., the product of NIC per item and the number of prescription items) on these products for each year. A graphical analyses of the HSCIC dataset will be found in Appendix 2B.

I used the information presented in Table 2.1, Appendices 2A and 2B, and the HSCIC dataset as the basis of the statistical modelling in the following sections of this chapter. Table 2.2 on page 32 describes *all* the variables used in my analyses; and which are constructed from the data sources listed above. The selection of these variables is based on the theoretical framework and arguments presented in section 2.2.2. The last paragraph of that section also provides a list of the variables of interest. A number of the variables are expressed in logged forms to account for possible skewness in the data and to enable the computation of elasticities with respect to a specified outcome or dependent variable[4].

[4] I found that the HSCIC dataset contained a multicollinearity feature: an approximately linear relationship between ln(N_PPTS) and the therapeutic class indicator D for the polio vaccines. An OLS regression of D on a set of all other explanatory variables that contained N_PPTS yielded an R-square of 1.000. I attribute this to the "single-product single-supplier" equilibrium created by the exclusive concentration of demand on IPV (see Appendices 2A and 2B). I dealt with this by dropping the variable ln(N_PPTS) from the regressions.

Table 2.1: A history of improvements in pharmaceutical manufacturing

Product Type	Process Innovation	Year of Introduction	Type of Process Innovation
Penicillins	Batch fermentation in lactose medium; sterilization performed in batches; use of depth filters for air filtration; bioassays; only temperature controlled for; batch-by-batch recovery and purification; mycelium removed by filtration; no recovery of side-chain precursors: it is discarded; tank reactor volumes of 10-20,000 gallons.	1950	Cost-reducing process innovation (titres of 0.5-1.0 g/L, efficiency of 70-80%, bulk product cost of US$275-350 per kg).
	Fed-batch fermentation in glucose/sucrose medium; continuous sterilization; use of membrane filters for air filtration; high-performance-liquid-chromatography assays; computerized control of the process (temperature, pH, dissolved oxygen etc.); semi-continuous recovery and purification; mycelium removed by whole-broth extraction; recovery of side-chain precursors for reuse; use of larger tank reactors (volumes of 20-60,000 gallons).	2000	Cost-reducing process innovation aimed at increasing productivity (titres of > 40g/L, efficiency > 90%, lower bulk product cost to US$15-20 per kg despite an increase in cost of labour, energy and raw materials) [32;33].
Hepatitis B vaccines	Plasma-derived hepatitis B vaccine licensed (after 13 years of research).	1981	-
	Recombinant hepatitis B vaccine licensed – the first viral subunit (virus-like particle) vaccine to prevent human cancer and first vaccine manufacturing using recombinant DNA technology. Production involves expression of hepatitis B surface antigen (HBsAg) in yeast Saccharomyces cerevisiae.	1986	Cost-increasing process innovation to address (1) the safety risks of using blood from previously infected people, and (2) the inadequate supply of human carrier plasma.
	Manufacturing improvements to reduce COG, allowing a reduction of price per dose to approximately $25.	1990's	Cost-reducing process innovation.
	Indian manufacturers express HBsAg in more productive yeast Pichia pastoris.	2006	Cost-reducing process innovation that allows supply of hundreds of million doses for a price of < 40¢ per vial [34-37].

MMR vaccines	Licensure of measles-mumps-rubella vaccine. The combination vaccine was introduced in the UK in 1988.	1971	Cost-increasing process innovation. Shift from separate monovalent vaccines to trivalent combination vaccines allows single dose administration in children [35].
Influenza vaccines	Licensure of purified influenza vaccine grown in embryonated chicken eggs.	1969,1970	-
	Adjuvanted killed (inactivated) influenza vaccine prepared in embryonated chicken eggs that is given by intramuscular or intradermal administration.	1973	-
	Switch from inactivated influenza vaccines to live attenuated influenza vaccines, also grown in fertilized chicken eggs.	2003	Cost-reducing process innovation that allowed intranasal (needle-free) delivery of the vaccine.
	Gradual but growing interest in moving from egg-based manufacturing to inactivated influenza vaccine grown in tissue cell culture, which could be Madin-Darby canine kidney cells; so-called VERO cells derived from African Green monkeys; or human embryonic retinal cells.	2004, 2006	Cost-increasing process innovation to increase manufacturing capacity, provide flexible and speedy manufacturing scale-up to meet fluctuations in demand – and ensure reliability of supply; and offer an alternative for people allergic to egg proteins [35;37].
Polio vaccines	Licensure of killed (inactivated) Salk polio vaccine grown in primary monkey kidney cells.	1955	-
	Licensure of oral (attenuated/weakened) Sabin polio vaccine grown in so-called VERO cells derived from monkey kidney cells.	1957	Cost-reducing process innovation: oral polio vaccine (OPV) was cheaper to make and cheaper to administer (than inactivated polio vaccine [IPV]). IPV required repeated booster doses whist OPV offered longer-lasting immunity.
	Development of improved/enhanced Salk IPV.	1960's through 1970's	Cost-increasing process innovation to address safety concerns about OPV risk of vaccine-associated paralytic polio and vaccine-derived polioviruses[36-39].

Table 2.2: List of variables

Variables	Definitions	Mean (SD)
COSTREDINNOV	Dummy indicator for products embodying a cost-reducing process innovation (= 1) or cost-increasing process innovation (= 0) as defined in Table 1. The variable refers to product innovations used by firms to make the medicines supplied as well as improvements made to the manufacturing process itself.	0.85 (0.35)
YRSINDEXP	Length of time (in years) of industrial experience in using a given process innovation prior to each cross-sectional wave of the dataset. The variable captures the time-dimension to process improvements, i.e., regardless of whether a process innovation (in comparison with an earlier process) is cost-reducing or not, the period after its introduction will be characterized by cumulative (year-after-year) efficiency gains from learning-by-doing effects and economies-of-scale.	45.92 (10.88)
COSTREDINNOV*YRSINDEXP	Interaction term of COSTREDINNOV and YRSINDEXP that captures any *joint* impacts of the nature of process innovation and the cumulative efficiency gains over time.	42.3 (18.33)
PRODUCTAGE	This refers to the time gap between a given cross-sectional wave and the year in which a product introduced on the NHS market or the year 1991 if there was no information on when a product was introduced on the NHS.	9.78 (6.35)
$PRODUCTAGE^2$	A quadratic term for PRODUCTAGE that is intended to: (1) capture any (skewed) inverted U-shaped distribution that might exist for product density over time; and (2) test whether $\partial P_{ijc}/\partial Age_{ijc} < 0$ is subject to diminishing marginal effects.	136.04 (153.49)
ln(N_FIRMS)	Natural logarithm of the number of firms or sellers in each therapeutic market. Controlling for the number of suppliers indirectly captures the impact of market-entry barriers due to trade secrets, process patents, intellectual property disputes, or difficulties in duplicating a process innovation.	3.30 (0.96)
IMPORTS	Dummy indicator for products imported. This takes the value of one if a product is imported and zero otherwise.	0.017 (0.13)
ln(N_PRODUCTS)	Natural logarithm of an aggregated measure of the number of products (in a given year).	4.60 (1.10)
ln(N_TS)	Natural logarithm of the number of therapeutic-substitute molecules (in a given year).	2.51 (0.95)
ln(N_PPTS)	Natural logarithm of the number of products per therapeutic-substitute molecule (in a given year).	1.91 (0.44)
ln(N_GE)	Natural logarithm of the number of generic equivalents for a given product (in a given year).	3.30 (0.96)
ln(DEMANDVOL)	Natural logarithm of demand volumes measured by the number of prescription items (in a given year).	6.77 (4.40)
ln(PRODUCTPRICE)	Natural logarithm of product prices measured by the net-ingredient-cost per item.	2.38 (1.40)
A	Dummy indicator for the penicillins as a group. This takes on the value of one and zero otherwise.	0.85 (0.36)
B	Dummy indicator for hepatitis B vaccines. This takes on the value of one and zero otherwise.	0.07 (0.25)
C	Dummy indicator for influenza vaccines. This takes on the value of one and zero otherwise.	0.06 (0.23)
D	Dummy indicator for polio vaccines. This takes on the value of one and zero otherwise.	0.01 (0.12)
E	Dummy indicator for MMR vaccines. This takes on the value of one and zero otherwise.	0.01 (0.12)

Notes: SD = standard deviation. The indicators of therapeutic-class or 'market' are intended to control for therapeutic-class-specific effects. I implement this in the regressions as a set of four dummy variables (A, B, C and D) with the MMR vaccines embodying cost-increasing process innovations and not imported as the baseline or reference category. This is because of the demand uncertainty created by the 'Wakefield Hoax', which I consider as a random historical event. Since I do not know of any similar random historical events or controversies that created demand uncertainty for the other product classes, I believe this should help control confounding and allow identification of the *ceteris paribus* effects of manufacturing process innovations.

2.3.3 Empirical Modelling

To test the hypotheses listed in section 2.3.1, I conducted some econometric modelling that is best appreciated if one considers that, in contrast to Figure 1, in the real world, the products in my sample will have in each year and over time, different NHS demand curves with discontinuities and undefined segments. This can be attributed to: disease epidemiology (for example, severity or rarity of an illness); uncertainty in medicinal demands; advertising, marketing and changing government regulations; entry of new products (for unmet clinical needs) and exit of other products compensated by importation of products from foreign suppliers when and where demand exceeds local supply; random historical events etc. That aside, within each year, the sample products supplied will embody the product innovations (cost reducing or not) used by manufacturers to make these products as well as the process improvements made up to that year.

This is best captured within a multilevel or longitudinal framework where there is natural clustering of products demanded by the NHS within each time period. Using the terminology of Rice and Jones [25], the individual products are micro-units clustered within each cross-sectional year as the macro-unit. For each hypothesis, I estimated a number of econometric models to check whether the outputs of the preferred model is consistent with alternative competing models, even if one believes the outputs of the preferred model are biased.

Hypothesis 1. With this, I wanted to establish the nature of the relationship between (cost-reducing) process innovation and product density. Using a multilevel framework, I specified a linear random-intercept model, in which the macro-units can be treated as having random-effects (RE) or fixed effects (FE). The RE model assumes error terms for the macro- and micro-units are randomly distributed and it can be written as:

$$\begin{aligned} Y_{it} &= \beta_0 + \boldsymbol{\beta_1}\mathbf{X}_{it} + \mu_{0t} + \epsilon_{it} \\ \mu_{0t} &\sim N\left(0, \sigma_\mu^2\right),\ \epsilon_{it} \sim N(0, \sigma_\epsilon^2) \\ VPC &= \sigma_\mu^2/\sigma_\mu^2 + \sigma_\epsilon^2\,, \qquad \mathrm{corr}(\mathbf{X}_{it}, w_{it}) = 0 \end{aligned} \tag{2}$$

where Y_{it} is the outcome of interest; $\mathbf{X}_{it}$ is a vector of explanatory variables, β_0 is the intercept and $\boldsymbol{\beta_1}$ is a vector of coefficients. The subscript *i* (= 1, 2, 3, …, M) indexes the individual products supplied by manufacturers; *t* (= 1, 2, 3, …, T) indexes the cross-sectional years of the dataset; $w_{it} = \mu_{0t} + \epsilon_{it}$ is the composite error term, where ϵ_{it} refers to

idiosyncratic errors that vary over individual products; μ_{0t} refers to year-specific unobservable error and VPC is an error variance partition coefficient.

The FE model, which considers the error term for the macro-unit as a fixed parameter to be estimated, can be written as:

$$\begin{gathered} Y_{it} = \beta_0 + \mu_{0t} + \boldsymbol{\beta_1 X}_{it} + \epsilon_{it} \Leftrightarrow \beta_{0t} + \boldsymbol{\beta_1 X}_{it} + \epsilon_{it} \\ \beta_{0t} = \beta_0 + \mu_{0t}, \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \text{corr}(\mathbf{X}_{it}, \mu_{0t}) \neq 0, \text{corr}(\mathbf{X}_{it}, \epsilon_{it}) = 0 \end{gathered} \tag{3}$$

The random-intercept models above can be modified into (1) a pooled or population-averaged [PA] model or (2) a mixed-effects [ME] linear model that allows consideration of both random-intercepts and random coefficients. The PA model is a pooled ordinary least squares estimator that can be written as follows:

$$\begin{gathered} Y_{it} = \beta_0 + \boldsymbol{\beta_1 X}_{it} + v_{it} (= \mu_{0t} + \epsilon_{it}) \\ \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \text{corr}(\mathbf{X}_{it}, v_{it}) = 0 \end{gathered} \tag{4}$$

where $\mu_{0t} = 0, \text{and } v_{it} = \epsilon_{it}$. If $\mu_{0t} \neq 0$, then $v_{it} \neq v_{is}, \text{for } t \neq s$ and the error terms will be serially correlated even if ϵ_{it} is identical and independently distributed with a zero mean and variance σ_ϵ^2. For the ME model, where random-intercepts and random-slopes can be considered simultaneously, we can write the following:

$$\begin{gathered} Y_{it} = \beta_0 + (\boldsymbol{\beta_1} + \mu_{1t})\mathbf{X}_{it} + \mu_{0t} + \epsilon_{it} \\ \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \begin{pmatrix} \mu_{0t} \\ \mu_{1t} \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Omega_\mu}), \mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Omega_\mu} = \begin{pmatrix} \sigma_{\mu 0}^2 & \\ \sigma_{\mu 01} & \sigma_{\mu 1}^2 \end{pmatrix} \end{gathered} \tag{5}$$

where μ_{1t} captures randomness in the coefficients (slopes), $\sigma_{\mu 0}^2$ is the variance of the random intercept, $\sigma_{\mu 1}^2$is variance of the (selected) random slopes and $\sigma_{\mu 01}$ is covariance between the random intercepts and random coefficients. I estimated the ME model with the variable for cost-reducing process innovation having a random slope [26; 27].

For product density as the outcome of interest, I estimated the RE, FE, PA and ME models using the following regressors from Table 2.2: COSTREDINNOV, YRSINDEXP, COSTREDINNOV*YRSINDEXP, PRODUCTAGE and its quadratic term; ln(N_FIRMS), IMPORTS, ln(DEMANDVOL), ln(PRODUCTPRICE) and the therapeutic-market indicators: A, B, C, D.

For the second hypothesis, I moved away from the multilevel framework to a "survivorship technique". I initially performed a non-parametric survival analysis under the assumption of non-informative censoring. I constructed Nelson Aalen plots of the cumulative hazard of product exit differentiated by (1) therapeutic classes; and (2) the cost-reducing nature of the process innovation(s) deployed in drug manufacturing. A well-known problem with such non-parametric survival analysis is that it doesn't allow for any differences observed to be confidently attributed to the key variable of interest, COSTREDINNOV. That is, one cannot tell whether other confounding variables contributed more or less to the differences in exit hazard. I therefore conducted a number of parametric survival analyses using: (1) continuous-time Cox, Weibull and Gompertz models, and (2) discrete-time complementary log-log models (hereafter cloglog).

In the Cox model, the exit hazard for product *i* at time *t*, given survival to that time, is defined by the product of a baseline hazard, $h_0(t)$ as an *unspecified* function of time, and a proportionality factor as a function of explanatory variables. This is as follows:

$$h_i(t|\mathbf{X}_i) = h_0(t).\exp(\mathbf{X}_i\boldsymbol{\beta_1} + \mu_{0t}) \quad (6)$$
$$h_i(t|\mathbf{X}_i)/h_0(t) = \exp(\mathbf{X}_i\boldsymbol{\beta_1} + \mu_{0t}) \Leftrightarrow \ln[h_i(t|\mathbf{X}_i)/h_0(t)] = \mathbf{X}_i\boldsymbol{\beta_1} + \mu_{0t}$$

where μ_{0t} is a gamma-distributed unobserved heterogeneity (frailty) that is shared by all products consumed in the NHS in each given year. This shared frailty has unique values across time periods but constant in each year. Depending on the parametric specification of the baseline hazard in relation to time, we can write the following:

$$\text{if } h_0(t) = hp.\,t^{\rho-1}, \quad (7)$$
$$\ln[\,h_i(t|\mathbf{X}_i)] = (\rho-1)\ln t + \beta_0 + \mathbf{X}_i\boldsymbol{\beta_1} + \mu_{0t} \Rightarrow \text{Weibull model;}$$

$$\text{if } h_0(t) = h.\exp(\gamma t),$$
$$\ln[\,h_i(t|\mathbf{X}_i)] = \gamma t + \beta_0 + \mathbf{X}_i\boldsymbol{\beta_1} + \mu_{0t} \Rightarrow \text{Gompertz model,}$$

for which ρ and γ are ancillary parameters to be estimated and β_0 is an intercept. If $\rho > 1$ for the Weibull model, the hazard of product exit is continuously increasing; if $\rho < 1$, the hazard is decreasing, and when $\rho = 1$, an exponential (constant hazard) model is obtained. If $\gamma > 0$ for the Gompertz model, the hazard of product exit is increasing monotonically; if $\gamma < 0$, the

hazard is decreasing and if $\gamma = 0$, a constant hazard is obtained. I estimated the continuous-time models with and without shared unobserved heterogeneity [27].

Yet the complexities of real-life are such that modelling the hazard of product exit over a continuous period may be inappropriate. Considering, for instance, uncertainties in NHS medicinal demands; discontinuity and time lags in developing and implementing manufacturing innovations (perhaps due to a reluctance to reopen the regulatory review process); it might be better to model the hazard of product exit over a series of short-run periods. I therefore partitioned the continuous time-period (in this case1991-2012) into a sequence of contiguous periods: $(0, t_1], (t_1, t_2], \ldots, (t_{j-1}, t_j]$ where *j* indexes each discrete time-interval. (The parenthesis indicates the beginning of each time interval and the square bracket indicates its end.) I looked at events over 3-year periods, i.e., I partitioned the data into seven discrete time-intervals with the last time-interval comprising of four years.

I evaluated the grouped survival data created using a cloglog model with and without frailty, in which the hazard of product exit (h_{it}) within a discrete time-interval (*t*) is given by:

$$h_{it} = 1 - \exp\{-\exp[h_0^*(t) + \boldsymbol{\beta_1}\mathbf{X}_{it} + \theta]\} \tag{8}$$

$$\log(-\log(1 - h_{it})) = h_0^*(t) + \boldsymbol{\beta_1}\mathbf{X}_{it} + \theta$$

$$h_0^*(t) = \log\left[\int_{t_{j-1}}^{t_j} h_0(t)dt\right]$$

$$\theta \coloneqq \log(\mu_{0t}^*) \sim N\left(0, \sigma_\mu^2\right);\ \log(\mu_{it}^*) \sim \text{gamma}(1, \sigma_\mu^2)$$

where the expression $-\log(1 - h_{it})$ is the discrete-time analogue of $h_i(t|\mathbf{X}_i)$ in the continuous-time models; θ is the unobserved heterogeneity (frailty) that is uncorrelated with the explanatory variables; $h_0^*(t)$ is a non-parametric baseline hazard that is constant within each discrete time-interval [27],[28].

For the cloglog model, I adopted three representations of unobserved heterogeneity. First is a shared normally-distributed frailty (μ_{0t}^*) having distinct values of heterogeneity across each year of each discrete-time interval. Second is *observation-level* unobserved heterogeneity (μ_{it}^*): individual-product-specific frailty that varies across time periods – and which I assumed follows a gamma distribution. Third is a discrete non-parametric finite-mixture frailty that is characterised by two homogenous subpopulations or mass points consisting of: (1) 'frail' products with a higher unobserved exit propensity; and (2) products

with a lower latent hazard of exit[5]. The two groups may or may not be equally mixed within the sample population but over time, 'frail' products exit first and more frequently. The group with a low latent exit-hazard can be considered as consisting of "medically-necessary" drugs offering substantial clinical benefits in routine practice – and highly preferred by healthcare providers and consumers (on the basis of usage experience). The frail group consists of all other products. The values of the two mass points: m1 for the 'frail' group (normalized to zero) and m2 for the 'low' exit-hazard group, conditional on the probabilities of products belonging to each subpopulation, are estimated with all other parameters of the cloglog regression.

For the parametric survival analyses above, I used the following explanatory variables from Table 2.2: COSTREDINNOV, YRSINDEXP and COSTREDINNOV*YRSINDEXP; PRODUCTAGE, IMPORTS, ln(N_FIRMS), ln(N_TS) and ln(N_GE); ln(DEMANDVOL), ln(PRODUCTPRICE) and the indicators of therapeutic class (A, B, C and D).

Hypothesis 3. With this, I sought to investigate the relationship (if any) between improvements in pharmaceutical manufacturing (process innovation) and the observed price trends for individual products in my sample. Specifically, I aim to identify whether $\partial \mathrm{P_{ijc}}/\partial \mathrm{N_c} < 0,\ \partial \mathrm{P_{ijc}}/\partial \mathrm{N_{gj}} < 0$ and $\partial \mathrm{P_{ijc}}/\partial \mathrm{costredinnov_{ijc}} < 0$. Going back to the multilevel framework, I first constructed a random intercept RE model using ln(PRODUCTPRICE) as the dependent variable – and the following explanatory variables: COSTREDINNOV, YRSINDEXP and COSTREDINNOV*YRSINDEXP; IMPORTS, PRODUCTAGE and its quadratic variant; ln(N_FIRMS), ln(N_TS) and ln(N_GE), ln(DEMANDVOL) and the dummies A, B, C and D for therapeutic class. I rerun the analysis using the FE[6], PA and ME models.

[5] In principle, the number of mass points should be chosen such that an additional mass point makes no improvement to the log-likelihood function and the prior probability of products belonging to that additional mass point is close to zero. I found that moving to three mass points resulted in more convergence problems. I therefore maintained the default of two mass points for the finite-mixture frailty.

[6] Note that because product prices are determined by both "supply" and "demand" factors, there might be an endogeneity (simultaneity) issue here. One solution is to separate out time-invariant and time-varying exogenous and endogenous variables, specify variables internal to the model to be used as instruments for the endogenous variables; and use the instrumental-variables estimators suggested by Hausman and Taylor (HT) and Amemiya and MaCurdy (AM)[27]. These estimators allow estimation of the $\widehat{\boldsymbol{\beta}}$ vector for time-invariant variables and are appropriate when the variables are correlated with the unobserved individual-product error ($\widehat{\epsilon_{\mathrm{it}}}$) but inconsistent when the variables are correlated with the unobserved period-specific error ($\widehat{\mu_{0\mathrm{t}}}$). Relative to the FE estimator, the majority of gains from the HT and AM estimators is with respect to $\widehat{\boldsymbol{\beta}}$ for the time-invariant variables. Since the HSCIC dataset has no time-invariant variables, these efficiency gains are not relevant here. In fact, running the –xthtaylor– command in STATA (with or without "amacurdy" option) and excluding time-invariant variables yields an error message.

To recap, my regressions on product density and product prices utilized fixed-effects, random-effects, population-averaged and mixed-effects estimators. The regressions on hazard of product exit makes use of continuous-time (Cox, Weibull, Gompertz) and discrete-time cloglog estimators. I performed all my analyses in STATA v.11.

2.4 Results

2.4.1 Product Density and Process Innovation

The full set of results from regressions on product density are presented in Appendix 2C. One observes a general consistency in the coefficients estimated from the FE, RE, PA and ME models without a random coefficient for the variable for cost-reducing process innovation. A standard Hausman test for a comparison of the FE and RE models suggested that there are statistically-significant differences between $\widehat{\beta_{FE}}$ and $\widehat{\beta_{RE}}$ that indicates a preference for the FE model. An alternative test based on the so-called augmented regression lends further support for the FE model. On the basis of the log-likelihoods and Akaike Information Criterion (AIC), selected model results are presented in Table 2.3 below. Since the log-likelihood of the FE model is greater than that of the ME model without a random slope for the variable for cost-reducing process innovation, and the corresponding AIC is also lower, I focus, momentarily, on the FE model.

The FE model shows that the process-innovation variables (COSTREDINNOV and YRSINDEXP) have generally a positive statistically-significant effect on product density. This positive effect is, however, counteracted by the negative coefficient on the interaction term for the variables for cost-reducing process innovation and years-of-industrial experience (COSTREDINNOV*YRSINDEXP). With years-of-industrial experience equated to five years[7], products that are manufactured with a cost-reducing process innovation have 2.76 (= 1 + [exp(1.1672 – 0.0303(5)) – 1]) times the density of products manufactured with cost-increasing process innovations, holding all else equal.

The coefficient for the variable for years-of-industrial experience indicates that, for each additional year of using a cost-increasing process innovation, the gains from learning-

[7] Within the HSCIC dataset, the variable years-of-industrial experience has a mean of approximately 46 years, a minimum of 5 years and a maximum of 62 years. I, in most cases, estimate the partial marginal effects of the variable for cost-reducing process innovation (over the long-run) by assuming 20 years of industrial experience. The "long run" here refers to the period 1991 to 2012.

by-doing and economies-of-scale contributes, on average, 0.93% increase in product density: a positive but small effect. For products manufactured with cost-reducing process innovations, product density after two decades is 75.28% (= 100•[exp(1.1672 – 0.0303(20)) – 1]) higher. That the effect of the variable for cost-reducing process innovation on product density declines with years-of-industrial experience is perhaps because of dynamic replacement of what becomes 'obsolete' manufacturing methods by newer ones; a more limited scope, or easily exhausted opportunities for process improvements; and/or that efforts to lower COG are not repeated year after year. Another possible explanation is: products embodying cost-reducing process innovations, being subject to more intense price-discounting competition (in overcrowded therapeutic markets), may have relatively higher exit hazards – see sections 2.4.2 and 2.4.3 below.

The FE model shows that variable for imported products has a positive effect on product density: compared to a dependence on only locally-sourced products, imports increase product density by approximately 5.37% (= 100•[exp(0.0523) – 1]. This is not a particularly surprising result as the primary motivation behind importation is to correct a perceived or real mismatch between drug supply and demand forecasts. Another unsurprising result is observed with the coefficient on the log of the number of manufacturing firms: a 10% increase in the number of firms is associated with a 7.51% increase in product density. On the other hand, demand volumes and product prices do not seem to have statistically-significant effects on product density – observe the near zero coefficients (elasticities) for the log of demand volumes and the log of product prices. Recall that the volume of quasi-rents available to be appropriated determine the incentives for firm *and* product entry. It thus appears that controlling for the number of manufacturing firms and therapeutic-class-specific effects, leaves a lesser amount of variation in product density that can be explained by demand volumes and product prices.

I now turn my attention to the results of the ME model with a random coefficient for the variable for cost-reducing process innovation – as the log-likelihood (AIC) of this model is greater (lower) than that for the FE model. This suggests that the effect of the variable for cost-reducing process innovation on product density is not fixed but varies across time periods (the macro-units). To be specific, the effects of the variable are lower (higher) than average in time periods where the reference category (i.e., MMR vaccines embodying cost-increasing process innovations and sourced locally), has higher (lower) product densities: a correlation coefficient of -0.9701. The average (short-run) effect of the variable for cost-reducing process innovation is to increase product density by 31.09% (= 100•[exp(0.2952 –

0.0049(5)) – 1]), holding all else equal. This heterogeneity in the variable's effects over the study period suggest that short-run adjustments differ from the average long-run effects of process innovations on product density.

Table 2.3: Regression analysis on product density

Dependent variable: ln(N_PRODUCTS)			Number of observations: 3721
Model specified:	Fixed-effects model	Mixed-effects linear model	
Coefficients:	$\hat{\beta}_{FE}$(SE)	$\hat{\beta}_{ME}$(SE)[a,c]	$\hat{\beta}_{ME}$(SE)[b,c]
COSTREDINNOV	1.1672 (0.421)*	1.1522 (0.059)***	0.2952 (0.086)**
YRSINDEXP	0.0093 (0.003)*	0.0100 (0.001)***	0.0008 (0.001)
COSTREDINNOV*YRSINDEXP	-0.0303 (0.008)**	-0.0300 (0.001)***	-0.0049 (0.002)**
PRODUCTAGE	-0.0008 (0.001)	-0.0008 (0.001)	0.0002 (0.001)
PRODUCTAGE^2	0.00001 (0.0004)	0.00001 (0.0004)	-0.00002 (0.0004)
ln(N_FIRMS)	0.7514 (0.058)***	0.7529 (0.013)***	0.6593 (0.014)***
IMPORTS	0.0523 (0.018)**	0.0524 (0.010)***	0.0552 (0.0093)***
ln(DEMANDVOL)	0.0001 (0.0001)	0.0001 (0.0003)	0.0002 (0.0003)
ln(PRODUCTPRICE)	-0.0003 (0.0005)	-0.0003 (0.001)	0.0001 (0.001)
A	2.0655 (0.240)***	2.0474 (0.063)***	2.0735 (0.060)***
B	1.2527 (0.106)***	1.2631 (0.021)***	1.1549 (0.020)***
C	0.5422 (0.089)***	0.5433 (0.019)***	0.6123 (0.018)***
D	0.7832 (0.101)***	0.7824 (0.021)***	0.6693 (0.020)***
Regression statistics			
Intercept: $\widehat{\beta_0}$(SE)	0.1238 (0.122)	0.1323 (0.049)**	0.4635 (0.066)***
$(\widehat{\sigma_{\mu 0}}, \widehat{\sigma_{\epsilon}}; \widehat{\sigma_{\mu 1}})$	(0.177, 0.078)	(0.167, 0.078)	(0.265, 0.070; 0.252)
Covar$(\widehat{\mu_{0t}}, \widehat{\mu_{1t}})$, corr$(\hat{\beta}_0, \hat{\beta}_{1t})$	–	–	-0.0648, -0.9701
Variance partition coefficient	0.8362	0.8213	0.4596
Overall R-squared	0.9769	–	–
Corr$(\mathbf{X}_{it}, \widehat{\mu_{0t}})$	-0.4833	–	–
Log-likelihood (AIC)	4221.6 (-8417.1)	4137.3 (-8242.6)	4510.4 (-8984.8)

Notes: SE = heteroskedastic-robust standard errors for FE, RE and PA models, default errors for the ME models; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; [a] Without a random-slope for COSTREDINNOV; [b] With a random-slope for COSTREDINNOV; [c] Log-likelihood ratio (LR) test for a comparison with the equivalent single-level model was statistically significant. LR test for a comparison of (a) and (b) indicates statistically-significant evidence that the effect of COSTREDINNOV is not fixed: it varies across different time periods; AIC = Akaike Information Criterion.

2.4.2 Exit Hazard and Process Innovation

The findings from my non-parametric survival analysis are depicted graphically in Figures 2.2 and 2.3. This shows the cumulative hazard of product exit over the time period 1991 to 2012, differentiated by: (1) the therapeutic market, and (2) the cost-reducing nature of the process innovation used in drug manufacturing. In general, one observes a low cumulative exit hazard (below 10%) for all therapeutic classes with the exception of the polio vaccines.

Figure 2.2 shows that the hepatitis B and influenza vaccines have the lowest exit hazards. In fact, no exit events are observed for these vaccines until after the year 1996. The exit hazard for the MMR vaccines first rises steadily before tapering off after year 2000, which is roughly the time of the outbreak of the MMR controversy. The penicillins exhibit a constant increase in the hazard of exit over time: a trend that is consistent with the inverted U-shaped distribution for product density and the shakeout period observed in Figure B1 in Appendix 2B. Consistent with the trend towards a "single-product single-supplier" equilibrium, the cumulative exit hazard for polio vaccines rises to just under 30% by year 2012. On the other hand, Figure 2.3 shows that exit hazard is generally higher for products embodying a cost-reducing process innovation relative to those made with a cost-increasing process innovation. A Wilcoxon test shows a statistically-significant difference in exit hazard (p value = 0.0003) and a log-rank test confirms this (p value = 0.0009).

Figure 2.2: Hazard of product exit by therapeutic class

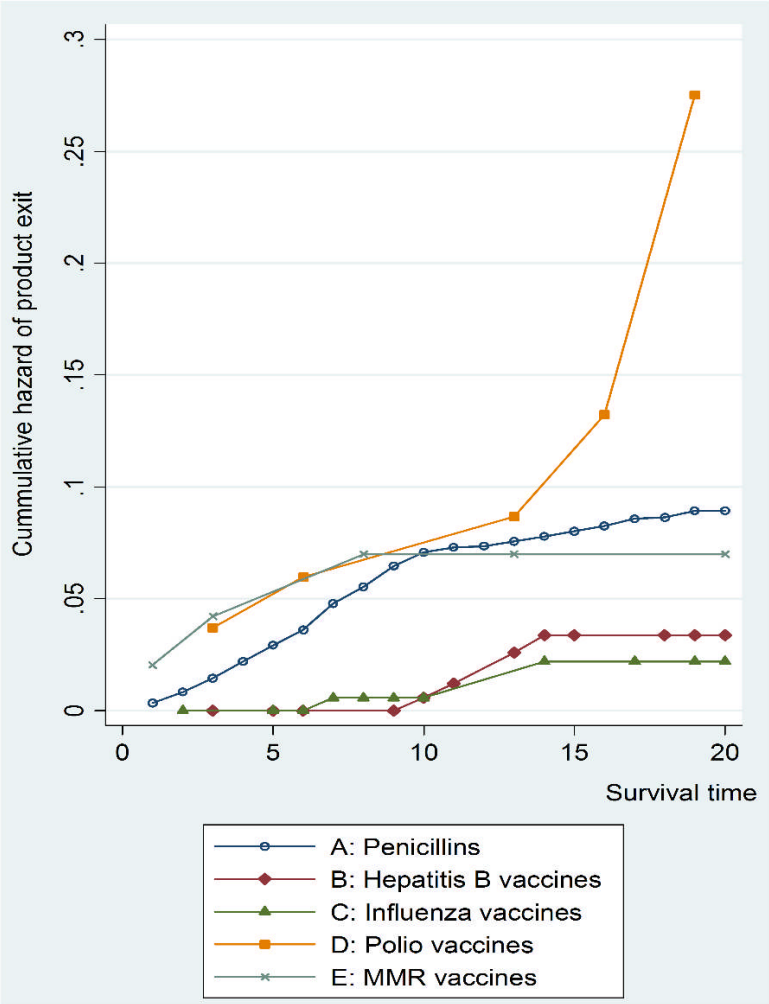


Figure 2.3: Hazard of product exit by type of process innovation

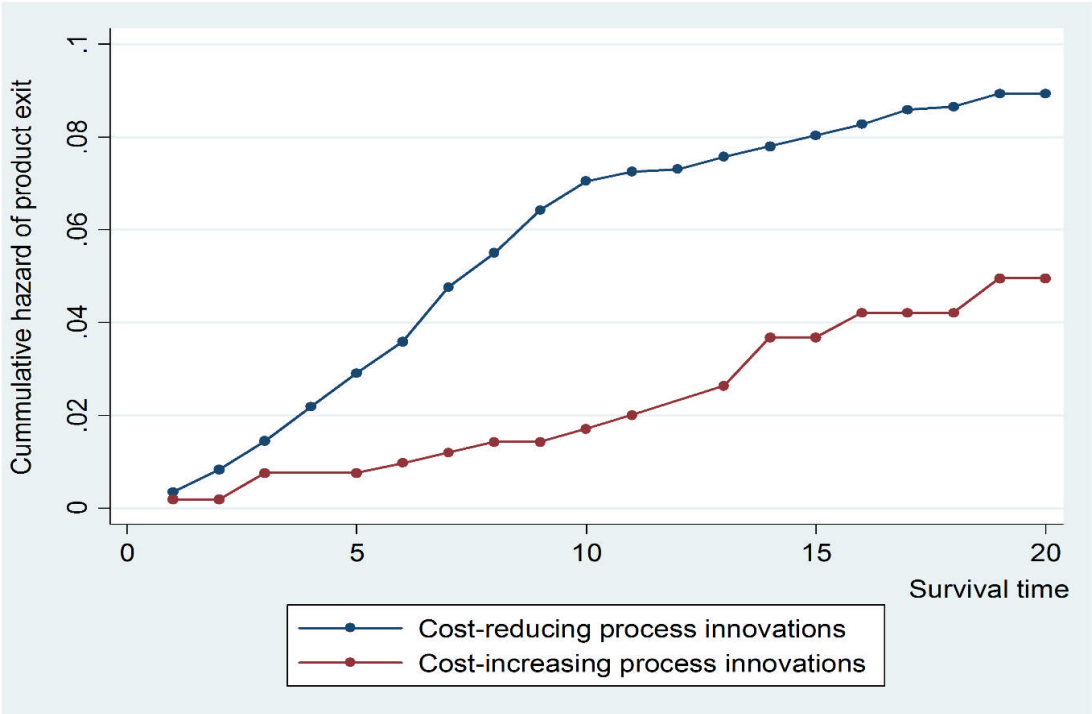


Table 2.4A: Regression analysis on hazard of product exit – without unobserved heterogeneity

Dependent variable: Product exit			Number of observations: 3673	
Model specified:	Weibull model		Cloglog model	
Coefficients:	$\hat{\beta}_1$(SE)	HR	$\hat{\beta}_1$(SE)	HR
COSTREDINNOV	-3.4433 (4.001)	0.0320	0.4095 (3.302)	1.5061
YRSINDEXP	0.0828 (0.064)	1.0863	0.0120 (0.037)	1.0121
COSTREDINNOV*YRSINDEXP	0.0532 (0.091)	1.0547	0.0454 (0.070)	1.0465
PRODUCTAGE	-0.0916 (0.019)***	0.9124	0.2363 (0.030)***	1.2665
ln(N_FIRMS)	3.1111 (0.652)***	22.4459	1.2796 (0.755)!	3.5951
ln(N_GE)	-0.2218 (0.070)**	0.80110	-0.0759 (0.073)	0.9270
ln(N_TS)	-1.8181 (0.580)**	0.1623	-2.1717 (0.668)**	0.1140
IMPORTS	-0.5315 (0.450)	0.5877	-0.3051 (0.444)	0.7371
ln(DEMANDVOL)	-0.6116 (0.042)***	0.5425	-0.5177 (0.058)***	0.5959
ln(PRODUCTPRICE)	-0.2680 (0.051)***	0.7649	-0.0631 (0.052)	0.9389
A	-3.2243 (2.491)	0.0398	-0.2805 (2.827)	0.7554
B	2.1866 (1.6464)	8.9045	1.8047 (1.067)!	6.0781
C	-2.3171 (1.232)!	0.0986	-0.2721 (1.352)	0.7618
D	4.2768 (0.884)***	72.0131	1.8534 (0.893)*	6.3816
(0,t0]	–	–	-2.2455 (1.516)	0.1059
(t0,t1]	–	–	-2.8780 (1.528)!	0.0562
(t1,t2]	–	–	-3.7825 (1.541)*	0.0228
(t2,t3]	–	–	-5.0872 (1.516)***	0.0062
(t3,t4]	–	–	-6.1063 (1.456)***	0.0022
(t4,t5]	–	–	-6.7779 (1.522)***	0.0011
(t5,t6]	–	–	-9.6622 (1.634)***	0.0001
Regression statistics				
Intercept: $\widehat{\beta_0}$(SE)	-8.4685 (2.259)***		–	
Ancillary parameter: ρ, γ (SE)	1.4842 (0.071)***		–	
Log-likelihood (AIC)	-689.7 (1411.3)		-406.7 (855.4)	

Notes: HR = hazard ratio = exp($\hat{\beta}_1$); SE = heteroskedastic-robust standard errors for the continuous-time models, default errors for the discrete-time model; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; (0,t0]…(t5,t6] = discrete-time-interval dummies; AIC = Akaike Information Criterion.

Table 2.4B: Regression analysis on product exit – controlling for unobserved heterogeneity

Dependent variable: Product exit		Number of observations: 3673
Model specified:	Weibull model	Cloglog model
Coefficients:	$\hat{\beta}_1$ (SE)	$\hat{\beta}_1$ (SE)[a]
COSTREDINNOV	-3.5224 (4.044)	-0.2204 (9.619)
YRSINDEXP	0.0861 (0.059)	-0.0826 (0.097)
COSTREDINNOV*YRSINDEXP	0.0515 (0.091)	0.1134 (0.152)
PRODUCTAGE	-0.0760 (0.020)***	0.2935 (0.111)**
ln(N_FIRMS)	2.7630 (0.772)***	3.1174 (1.391)*
ln(N_GE)	-0.2301 (0.065)***	-0.0892 (0.208)
ln(N_TS)	-1.4200 (0.761)!	-4.4599 (1.601)**
IMPORTS	-0.5309 (0.405)	-0.5967 (1.271)
ln(DEMANDVOL)	-0.6139 (0.039)***	-0.6968 (0.134)***
ln(PRODUCTPRICE)	-0.2738 (0.044)***	0.1354 (0.117)
A	-3.1438 (3.048)	-0.3587 (9.247)
B	2.2799 (1.276)!	2.5370 (2.568)
C	-1.8953 (1.242)	-3.8139 (4.566)
D	3.9973 (1.023)***	2.9925 (2.437)
(0,t0]	–	-0.3862 (3.606)
(t0,t1]	–	-1.8735 (3.572)
(t1,t2]	–	-4.1532 (3.498)
(t2,t3]	–	-5.9550 (3.536)!
(t3,t4]	–	-8.2742 (3.638)*
(t4,t5]	–	-7.6864 (3.688)*
(t5,t6]	–	-12.2111 (5.634)*
Regression statistics		
Intercept: $\widehat{\beta_0}$ (SE)	-8.3810 (1.982)***	–
Log-likelihood (AIC)	-688.7 (1411.4)	-162.2 (370.3)
σ_μ^2 (Chi-square $\forall$ H_0: $\{\mu_{0t}; \mu_{it}\} = 0$)	0.0534 (1.91)!	–

Notes: SE = heteroskedastic-robust standard errors for the continuous-time model, default errors for the discrete-time models; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; [a] Model with finite-mixture unobserved heterogeneity; (0,t0]…(t5,t6] = discrete-time-interval dummies; H_0 = null hypothesis for unobserved heterogeneity; AIC = Akaike Information Criterion.

The full set of results from my parametric survival analysis without controlling for unobserved heterogeneity are shown in Appendix 2D. On comparing the log-likelihoods and AIC of the different models, I focus on the continuous-time Weibull and the discrete-time cloglog models; see Table 2.4A above. But bear in mind that the continuous-time Weibull model estimates exit hazard over a truncated "long-run" period (in this case 1991-2012) whilst the discrete-time cloglog model estimates exit hazard over a series of "short-run" periods indexed by time-interval dummies. The Weibull model shows that products made with cost-reducing process innovations have a lower exit hazard albeit this is not a statistically-significant effect. This yields a statistically-insignificant hazard ratio of 0.0926 (= exp(-3.4433 + 0.0532(20)) although this effect is not statistically significant. The ancillary Weibull parameter ($\rho = 1.4842$) suggests that baseline exit hazard increases monotonically over time. On the other hand, the cloglog model tells a slightly different story, which, of course, is dependent on how the survival data was partitioned. Within each short-run (i.e., each discrete time-interval), the coefficient of the variable for cost-reducing process innovation suggest an increase in exit hazard: a hazard ratio of 1.8899 (= exp(0.4095 + 0.0454(5)) albeit this is not statistically significant. This positive effect, however, is counteracted by the negative coefficients on the interval-specific dummies, which suggests a monotonic decline in the baseline exit hazard.

However, without controlling for unobserved heterogeneity, the models in Table 2.4A implicitly assume that all variation in exit hazard is determined by the chosen set of explanatory variables. This could lead to biased estimates of "duration dependence" (how exit hazards vary over time) and the incremental/ marginal effects of the explanatory variables specified. For this reason, I shift my focus to models that account for unobserved, omitted or unmeasured heterogeneity. A full set of results from these models are given in Appendix 2E. On the basis of the log-likelihood and AIC, I focus on the continuous-time Weibull model with shared frailty and the cloglog model with a two-point finite-mixture frailty; see Table 2.4B.

The Weibull model with frailty does confirm the negative impact of the variable for cost-reducing process innovation on long-run hazard of exit. Products embodying cost-reducing process innovations are less likely to exit: a 91.73% (= 100•[exp(-3.5224 + 0.0515(20)) – 1]) reduction – albeit statistically we are less confident that this is no different from zero. Not surprising, the coefficient on product age indicates that the older a product gets (by a year), the lower the hazard of exit: a 7.6% reduction. This perhaps indicates strong clinical preferences for ageing products that offer positive net health benefits relative to

newer drugs introduced on the NHS over the study period. The results also show a 10% increase in demand volumes is associated with 6.14% reduction in exit hazard; a 10% increase in product price is associated with 2.74% reduction in hazard of product exit; and the elasticity of exit hazard with respect to the number of firms is 2.76. The ancillary Weibull parameter in this case is slightly higher (1.4907).

The cloglog model with a two-point finite-mixture frailty suggests that the baseline hazard declines over each successive short-run period albeit this decline is not monotonic[8]. The model does confirm that products made with cost-reducing process innovations, in the short-run, have a higher exit hazard but again this is not statistically significant – the average effect of the variable for cost-reducing process innovation is approximately 41.43% (= 100•[exp(-0.2204 + 0.1134(5)) – 1]). The cloglog model with finite-mixture frailty also indicated, within each short-run: (1) higher exit hazard with increasing number of manufacturing firms; (2) increased hazard of exit with product age [due perhaps to the displacement of less-effective older drugs by newer more-effective ones]; (3) lower exit hazard with the number of therapeutic substitutes; and (4) no statistically-significant effects of product price and number of generic equivalents on exit hazard. Note that although the coefficient on the log of demand volumes is consistent with the findings of Virabhak and Sohn[29], the coefficients on the log of number of therapeutic substitutes and the log of the number of generic equivalents appear to be inconsistent. A plausible explanation is: I controlled for the number of manufacturing firms, which is associated with an increased hazard of product exit. The reported coefficients therefore indicate the marginal effects of the number of therapeutic substitutes and generic equivalents *per firm*, holding all else equal.

Table 2.4B shows shared frailty in the Weibull model did not reach statistical significance at the 5% level; i.e., results with and without frailty are not markedly different. In contrast, unobserved heterogeneity is statistically significant in the discrete-time cloglog model with finite-mixture frailty. The unobserved heterogeneity parameter $\widehat{m2}$ = -3.9205 (*p* value = 0.041) and there is a 0.59 probability that a given product belongs to the class with a low latent hazard of exit and a 0.41 probability of belonging to the class comprising of frail products. Conditional on class membership, the unobserved subpopulation of highly-preferred medically-necessary drugs are approximately 98% (= 100•[exp(-3.9205) – 1]) less

[8] In an unreported analysis using the cloglog model, I specified in place of the discrete-time-interval dummies, a parametric baseline hazard of the form: $h_0 = (q - 1).\ln(t)$ where q is the discrete-time analogue of the ancillary parameter in the continuous-time Weibull model. The coefficient on ln(t) was -1.3249 (q = -0.3249) in the model without frailty, and in the model with gamma-distributed frailty this was -6.712 (q = -5.712). This confirms that, when the data is viewed as a series of short-run periods, the baseline hazard generally declines over time.

likely to exit, holding all else equal. The results of the discrete-time cloglog model with finite-mixture frailty are thus preferred to the estimates reported in Table 2.4A.

Yet, I found, for HSCIC dataset, that the STATA estimators for the cloglog models with frailty were 'fragile', i.e., the log-likelihood as a function of the beta-coefficients is not globally concave and the results obtained may just be one of several local optima (a single unique global optimum is not guaranteed). It is safer then to view the estimates in Table 4B as not the 'final' or true values although the models reported no convergence issues at final iteration. What can be said is: irrespective of when one views 1991-2012 as a truncated representation of the "long-run" (in the case of the continuous-time Weibull model), or as a series of "short-run" periods (in the case of the discrete-time cloglog model); there is no strong evidence of any difference in the exit hazard for products embodying cost-reducing process innovations versus those made with cost-increasing process innovations.

2.4.3 Price Variation and Process Innovation

The full set of results from regressions on individual product prices are presented in Appendix 2F. Here one also observes consistent estimates from the FE, RE, PA and ME models. A standard Hausman test for a comparison of the FE and RE models suggested that there are statistically-significant differences between $\widehat{\beta_{\mathrm{FE}}}$ and $\widehat{\beta_{\mathrm{RE}}}$ (p value < 0.0182) with a preference for the FE model. The alternative test based on the so-called augmented regression suggested that the difference between $\widehat{\beta_{\mathrm{FE}}}$ and $\widehat{\beta_{\mathrm{RE}}}$ are borderline statistically-significant (p value = 0.0476): a confirmation that the FE estimates are preferable. On the basis of the log-likelihoods and AIC, selected model results are presented in Table 2,5 below. Since the log-likelihood of the FE model is greater than that of the ME model with and without a random slope for the variable for cost-reducing process innovation; and the corresponding AIC is also lower, I focus on the FE model.

Table 2.5: Regression analysis on price variation (using individual product prices)

Dependent variable: ln(PRODUCTPRICE)			Number of observations: 3721
Model specified:	Fixed-effects model	Mixed-effects linear model	
Coefficients:	$\hat{\beta}_{FE}$(SE)	$\hat{\beta}_{ME}$(SE)[a,c]	$\hat{\beta}_{ME}$(SE)[b,c]
COSTREDINNOV	-3.1883 (1.131)*	-3.3181 (0.740)***	-3.3181 (0.740)***
YRSINDEXP	0.0119 (0.010)	-0.0212 (0.009)*	-0.0212 (0.009)*
COSTREDINNOV*YRSINDEXP	0.0373 (0.019)!	0.0465 (0.015)**	0.0465 (0.015)**
PRODUCTAGE	-0.0112 (0.018)	-0.0106 (0.012)	-0.0106 (0.012)
PRODUCTAGE^2	-0.0003 (0.001)	-0.0003 (0.001)	-0.0003 (0.001)
ln(N_FIRMS)	0.2174 (0.128)	0.2718 (0.137)*	0.2718 (0.137)*
ln(N_GE)	-0.4769 (0.025)***	-0.4748 (0.021)***	-0.4748 (0.021)***
ln(N_TS)	-0.1342 (0.098)	-0.0065 (0.154)	-0.0065 (0.154)
IMPORTS	0.9624 (0.159)***	0.9423 (0.162)***	0.9423 (0.162)***
ln(DEMANDVOL)	-0.0579 (0.006)***	-0.0581 (0.005)***	-0.0581 (0.005)***
A	1.4620 (0.591)*	1.2916 (0.731)*	1.2916 (0.731)*
B	1.4906 (0.202)***	0.8582 (0.239)***	0.8582 (0.239)***
C	0.4099 (0.196)*	0.1958 (0.327)	0.1958 (0.327)
D	0.3277 (0.196)	0.2970 (0.327)	0.2970 (0.327)
Regression statistics			
Intercept $\widehat{\beta_0}$(SE)	3.0685 (0.354)***	4.0042 (0.327)***	4.0042 (0.327)***
$(\widehat{\sigma_{\mu 0}}, \widehat{\sigma_{\epsilon}}; \widehat{\sigma_{\mu 1}})$	(0.249, 1.224)	(4.97E-9, 1.222)	(1.06E-8, 1.222; 9.59E-9)
Covar$(\widehat{\mu_{0t}}, \widehat{\mu_{1t}})$, corr$(\hat{\beta}_0, \hat{\beta}_{1t})$	–	–	< 0.0000001, 0.9088
Variance partition coefficient	0.0398	< 0.0005	< 0.0000001
Overall R-squared	0.2140	–	–
Corr$(\mathbf{X}_{it}, \widehat{\mu_{0t}})$	-0.2691	–	–
Log-likelihood (AIC)	-6013.1 (12054.2)	-6025.9 (12085.7)	-6025.9 (12081.7)

Notes: SE = heteroskedastic-robust standard errors for the FE, RE and PA models, default errors for the ME models; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; [a] Without a random-slope for COSTREDINNOV; [b] With a random-slope for COSTREDINNOV; [c] Log-likelihood ratio test for a comparison with the equivalent single-level model was not statistically significant. LR test for a comparison of (a) and (b) indicates no statistically-significant evidence that the effect of COSTREDINNOV is not fixed: it does not vary across different time periods; AIC = Akaike Information Criterion.

The FE model suggests that, over time, products that are manufactured with cost-reducing process innovations have significantly lower prices than those manufactured with cost-increasing process innovations. Ignoring the average time trend ($\widehat{\beta_0}$) that suggest increasing prices over the study period, the *incremental* effect of the variable for cost-reducing process innovation is a reduction in product prices of 91.30% (= 100•[exp(-3.1883 + 0.0373(20)) – 1]), holding all else equal. That is, the average drop over time in prices of products that are cheaper to make is roughly twice that of products that are costly to make. The coefficient on years-of-industrial experience and its interaction term (COSTREDINNOV*YRSINDEXP) all are positive but fail to reach statistical significance at the 5% level. That the marginal effects of years-of-industrial experience are "inconclusive" suggest that what matters more is whether clinically-beneficial drug products are manufactured with cost-reducing production plans, rather than how long firms have used a specific process innovation for drug manufacturing. On the other hand, this could be due to discontinuity in innovation efforts and/or time lags in developing and implementing process innovations.

The coefficient on the variable for imported products suggests that imported products, on average, carry a price premium of roughly 1.62 (= exp(0.9624) – 1) relative to locally-sourced products. This increase in payments probably reflects the fact that imported products are often sourced in emergency situations (to correct shortfalls between supply and demand), leaving little scope for healthcare payers to credibly exercise the threat of taking their business elsewhere. Holding all else equal, a 10% increase in the number for firms is associated with 2.17% increase in individual product prices. (Note that this effect is not significant in the FE model but significant in RE, PA and ME models.) This perhaps is because manufacturing firms seek to avoid any cutthroat price competition and engage in oligopolistic (imperfect) rivalry supplying new specialized or differentiated products for which they could charge a price premium.

I found that the coefficients (elasticities) for the variables product age, the quadratic for product age, log of the number of therapeutic substitutes, generic equivalents an demand volumes have the expected signs (as highlighted in section 2.2.2). However, for product age, the quadratic term for product age and the number of therapeutic substitutes, one is less confident (statistically) that their effects are not different from zero. Consistent with the notion that generic equivalents are closer treatment alternatives, and therapeutic substitutes compete on both quality and price dimensions, the *magnitude* of price reduction in response to the number of generic equivalents is greater than that due to the number of therapeutic

substitutes. A 10% increase in the number of generic equivalents results in a 4.77% price reduction whilst a 10% increase in the number of therapeutic substitutes results in 1.34% reduction in price. In addition, a 10% increase in demand volumes is associated with 0.58% reduction in individual product prices, holding all else equal. This is consistent with arguments by Ellison and Snyder [30] that the effect of purchaser size (demand volumes *per se*) on the magnitude of price discounts is small relative to the ability to engage in 'comparison shopping'.

Note, however, that the results of the ME model with a random coefficient for the variable for cost-reducing process innovation shows that there is no significant variation in the effect of the nature of process innovation used by manufacturing firms on individual product prices; the log-likelihood and AIC for the two variants of the ME model are roughly the same. This homogeneity in the slope coefficient for the variable suggest that, in contrast to the results for product density, short-run adjustments are no different from the average long-run effects of process innovation on individual product prices.

2.5 Discussion

Summing up the results, I want to emphasize two things. First, the preferred FE models use longitudinal variation in product density, product price; the variable for cost-reducing process innovation and other independent variables within individual products over time. As noted by Kaufman [31], this effectively ignores cross-section variation across the products within each given year. Granted, the FE model is generally consistent with the RE, PA and ME models that use more than just within-product variation.

Second, the relative effect of the variable for cost-reducing process innovation, on its own, is of little *practical* significance – and it applies to a baseline with cost-increasing process innovation. Take, for example, product prices as the outcome of interest. The best way to interpret my results is that in an environment where manufacturing firms will prefer to engage in oligopolistic rivalry (or insure against profit risk by introducing differentiated and specialized products carrying a price premium), socially-beneficial competition will be facilitated by the existence of price-sensitive healthcare payers (willing to offer incremental demand volumes in return for price cuts); a higher number of generic equivalents/ therapeutic substitutes and manufacturing products using cost-reducing process innovations. The results indicate the average drop, over time, in prices of products that are cheaper to make (for e.g. egg-based live-attenuated influenza vaccines) may be roughly twice that of products that are

costly to make (for e.g. tissue-cell-cultured influenza vaccines). This suggests that using cost-reducing process innovations may make firms more willing to consider discounted prices as a profit-maximizing strategy in competitive environments. Placed in the context of the work by Vernon et al. [4], my finding that products that are easy or less-costly to make also have lower prices fits more with boundary scenario *A* where some or all of the cost savings from manufacturing process improvements are passed on to consumers (via their healthcare payers) in the form of lower prices.

Besides the impact on price variation over time, I found that products manufactured with cost-reducing process innovations have, on average, higher product densities over time. This is consistent with theoretical predictions that cost-reducing process innovations effectively increase the volume of quasi-rents available to be appropriated and needed to sustain a multiplicity of manufacturing firms and a multiplicity of products. My findings also lend support to the notion that through improvements in manufacturing, the current cohort of biologics may go through the transition from difficult-to-manufacture/high-price/low-density products to easy-to-make/low-price/high-density products. My results, in addition, provide identifiers as to the type of products that are prone to risk of supply interruptions; to be precise, older products that are relatively expensive and complex to manufacture (or difficult to formulate) and have a small number of manufacturers. But I did not find any conclusive evidence that products embodying cost-reducing process innovations have lower instantaneous hazard of exit relative to products that are costly and difficult to manufacture.

There are, however, a number of issues with generalizing my findings to the current existing and emerging cohorts of biologic drugs.

The first relates to the fact that my product sample has a significant proportion of penicillins (> 80%) that are semi-synthetic small-molecule pharmaceuticals that are not made with purely bioprocessing methods – some chemical modification steps are involved. Why then did I include the penicillins in my sample? Why not just focus on the vaccines? I included the penicillins because they are thought to have gone through the transition from products that are difficult-to-manufacture and expensive-to-buy to commodity products that are cheap and easy-to-manufacture. Yet another reason for including the penicillins is that the principles behind their method of manufacturing have been used for the production of biologically-derived products including anticancer agents, hypercholesterolaemic agents and immunosuppressants [32]. What is more, the dummies A, B, C and D effectively controls for any unobserved or unmeasured class-specific effects. It is worth mentioning that the penicillins' richer history of manufacturing-cost reductions is partly due to changes that

occurred during and after World War II where public health urgency meant less stringent regulatory standards operated at the time [33]. The reported relative effects of the variable for cost-reducing process innovation may therefore be applicable to current and emerging biologics if and only if these products embody repeated efforts to reduce manufacturing costs over long periods of time.

Second, the manner in which I defined products as embodying cost-reducing or cost-increasing process innovations says nothing or, more accurately, hides the extent to which a given process improvement or innovation contributed to a fall or rise in COG. I am therefore unable to isolate the effects of incremental cost-reducing process innovations used to manufacture a given product or class of products. As discussed in section 2.3.1, only a direct approach (given availability of detailed data on COG for whatever product sample is selected) will provide answers to whether the magnitude of the cost-reduction achieved matters. This cannot be determined with my data. Third, because my data covers different periods of the lifecycle of products in my sample, my estimates may simply reflect longitudinal variations in product density, exit hazards, and prices at different points of market "disequilibrium". For instance, it is possible that the reported coefficient for the variable for cost-reducing process innovation may reflect the fact that most of the products in my sample are off-patent. Off-patent generic prices, under intense competition, depend more on manufacturing costs; whilst prices of branded on-patent drugs are largely determined by demand (for the hedonic/ quality attributes of these products). Hence a product sample with a lot more on-patent drugs may yield a coefficient that is perhaps closer to zero. This, however, is precisely what the framework in section 2.2.2 predicts: the outcome will be akin to that of the 'progressive' monopolist.

Fourth, using net-ingredient-cost-per-item as a proxy for product prices may have biased upwards the magnitude of the coefficients reported. Although my regressions controlled for demand volumes (proxied by the number of prescription items), this may not have fully compensated for the heterogeneity bias introduced by differences in pack sizes, dosages forms, dose strengths and duration of treatment. Finally, my analyses are based solely on data from the NHS; I did not consider drug consumption patterns in the UK private healthcare sector.

Based on these limitations, I suggest future research using a (more balanced) sample of products from the current and emerging cohorts of biologics to shed further insights. Notwithstanding, this research sheds light on a less-researched topic of the importance of manufacturing innovations in biologic-drug markets. The evidence here (i.e., a combination

of lower prices, higher product density, and no conclusive difference in hazard of product exit) suggests that repeated efforts to improve manufacturing efficiency could offer some insurance protection against drug shortages and reduce the dependence on emergency purchases of imported products. Even if supply interruptions or shortages occur, one will expect such interruptions to be brief and short-lived for products that are cheaper and easier to make. What is more, multiple product options should offer the scope to cater for heterogeneous clinical needs of patients. This carries important policy implications as to the sort of measures that can be used to address supply shortages, lower the budget impact of health technologies, and ensure more cost-effective delivery of health benefits to patients. This chapter argues that, compared to a reliance on emergency purchases of pricey imported products to increase product density (availability), a better solution is investing in bioprocess development (pharmaceutical manufacturing research).

2.6 Conclusions

There are substantial potential benefits from improvements in pharmaceutical manufacturing. Controlling for various sources of confounding, this study finds that biologically-derived pharmaceuticals made with cost-reducing process innovations may be associated with lower hazards of exit although I did not find any statistically strong evidence in support of this. I found statistically-significant evidence that products that embody cost-reducing process innovations tend to have higher product density and lower prices over time. The evidence suggests that making biologically-derived pharmaceuticals with cost-reducing process innovations should make manufacturers more willing to offer discounted prices as their profit-maximizing strategy when faced with price-sensitive healthcare payers. Based on my findings, one will expect the current cohorts of biologics (such as monoclonal antibodies) to make the transition from speciality medicines that are expensive and difficult-to-make to products that are cheap and easy-to-make – as long as these products embody cost-reducing process innovations and a history of repeated process improvements.

I will argue that investments in pharmaceutical manufacturing research (in developing cost-reducing process innovations) should be considered as one of the tools for preventing shortages of medically-necessary biologics and for encouraging socially-beneficial price competition to improve access, affordability and even more cost-effective delivery of health benefits to patients.

MANUFACTURING AND DRUG ADMINISTRATION COSTS

3 Systematic Review of Drug Administration Costs

3.1 Introduction

Increasingly healthcare payers are concerned about the affordability of biologics and question whether these technologies offer value for money. Sikora[53] and Kelly and Mirr[54], for instance, note that the rapid uptake of biologics for cancer, costing £30,000 per patient per year in the case of bortezomib (Velcade®), and £70,000 per patient per year for bevacizumab (Avastin®), cannot be sustained in publicly funded healthcare systems such as the UK NHS. While acquisition costs of biologics are considered to be relatively high compared with other drugs, the costs of treating patients with these drugs also depends on the cost of administration, which can vary substantially, and can be sizeable. It is feasible that a low-priced drug could have the same delivery costs as a high-priced drug once the administration costs are included. For example, two biologics *A* and *B* with acquisition costs of £7889 and £9980 per treatment cycle respectively will have roughly the same total delivery costs per cycle if biologic *A*, requiring 15 clinic visits or injections, incurs £3195 in administration costs whilst biologic *B*, requiring 5 clinic visits or injections, incurs £1065 in administration costs.

For this reason, when assessing value for money it is important, first, that administration costs are accounted for, and second, that the full range of these costs are included. There are, obviously, other cost components in an economic evaluation. The value for money offered by a biologic product, relative to other treatment options, is typically measured by the incremental cost-effectiveness ratio, which compares the discounted total (lifetime) costs of delivering the biologic with the discounted total (lifetime) health effects it is expected to offer. The discounted (lifetime) costs of each option is estimated by summing acquisition costs determined by the selling price; the associated administration costs determined by the chosen dosage regimen, route of administration and formulation; and future healthcare-related costs that are, in part, determined by the clinical effectiveness of that option. In this chapter, I focus on the resources costs of drug administration and how these costs vary from one dosage form to the other.

Following Aulton[55], I categorize routes of drug administration and associated dosage forms as: (1) oral routes involving solutions, syrups, tablets and capsules; (2) rectal routes involving suppositories and ointments; (3) topical routes involving the use of ointments, creams, topical aerosols and transdermal patches; (4) parenteral routes involving intravenous (i.v.), subcutaneous (s.c.) and/or intramuscular (i.m.) injections; (5) respiratory

routes involving the use of aerosols, inhalations and sprays; (6) nasal routes involving solutions and inhalations; (7) conjuctival and intraocular routes involving eye solutions, ointments and creams; and (8) solutions, suspensions, ointments and creams administered via the ear. According to Speers and Bonnano[56], oral formulations are the most preferred by patients and in terms of dosage regimen (frequency of administration), once per day dosing is most attractive to clinicians and patients. The authors define a hierarchy of drug (re)formulation that ranks modes of administration in the following order of increasing 'attractiveness': i.v. injections, i.m. injections, s.c. injections, injectable sustained release depot; transdermal (topical) applications; nasal and aerosol (respiratory) delivery; oral tablets or capsules administered more than once daily; and oral sustained release formulations that are administered once daily.

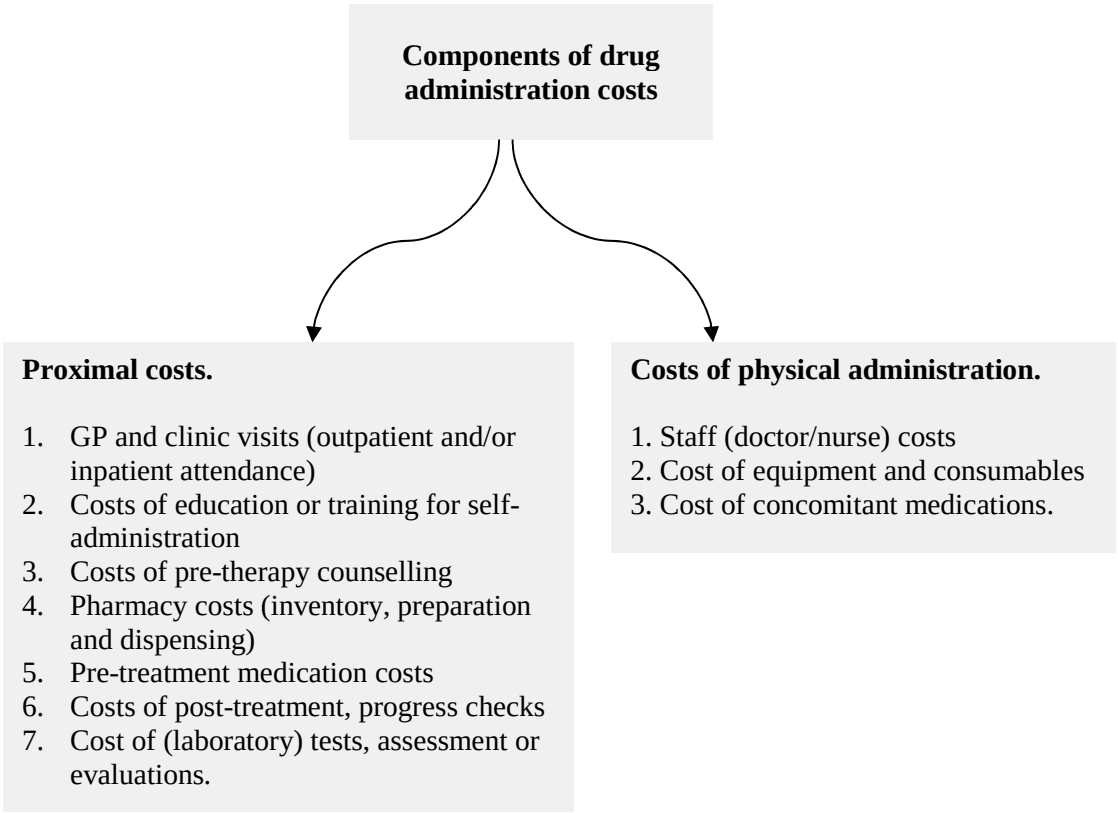
For biologics, consideration of the most appropriate drug delivery system is critical given they consist of large molecules with delicate structural forms of poor oral bioavailability. While oral biologics presents an opportunity for the industry in terms of increase in market demand and sales (Speers and Bonnano[56]), they also present a significant challenge for biopharmaceutical manufacturing as the delicate link between the macromolecular structure of biologics and their efficacy and safety profiles is such that moving from injectable to oral dosage forms may change the pharmacokinetic and/or pharmacodynamic properties. Although needle-free and oral delivery of biopharmaceuticals seem a desirable focus for manufacturers, parenteral, injectable routes seem to be the most realistic routes for most biologics, at least in the short term. An investigation of biologic administration costs associated with injectable routes is needed.

The aims of this chapter are to develop a framework of drug administration costs and, as a case example, use this framework to evaluate the administration costs of biologics within the UK NHS. My objective is not compare the magnitude of the costs of acquiring and administering drugs. As illustrated above, even if drug acquisition costs are greater than administration costs, differences (variation) in administration costs may be the determinant of differences in healthcare delivery costs. The objective here is (1) to conduct a systematic review of published, peer-reviewed studies to provide empirical estimates of the costs of administering biologic drugs within the UK NHS; and (2) to explore the sources of variation in these estimates, for instance, whether this is due to differences in the method of measurement or routes of administration. From the outputs of the literature review, I highlight the implications of my findings for research in biopharmaceutical manufacturing,

e.g., how healthcare delivery costs and associated drug formulation issues may affect or influence manufacturing and R&D decisions.

3.2 A framework of drug administration costs

Figure 3.1: Framework of drug administration costs



For this review, I adopted a broad definition of the components of administration costs covering not just the physical administration of a biologic drug through one of the routes mentioned above, but also *proximal* costs incurred before or after physical administration of a drug. Figure 3.1 above depicts a framework that is intended to highlight the need to be explicit about what constitutes drug administration costs. The framework serves as a guide for thinking through all potential costs associated with the administration of biologics and other medical technologies. The framework is generic as one would not expect the components of drug administration costs for biologics given via a particular route to differ markedly from that for non-biologics administered via the same route.

3.3 Literature review methods

3.3.1 Search strategy and study selection

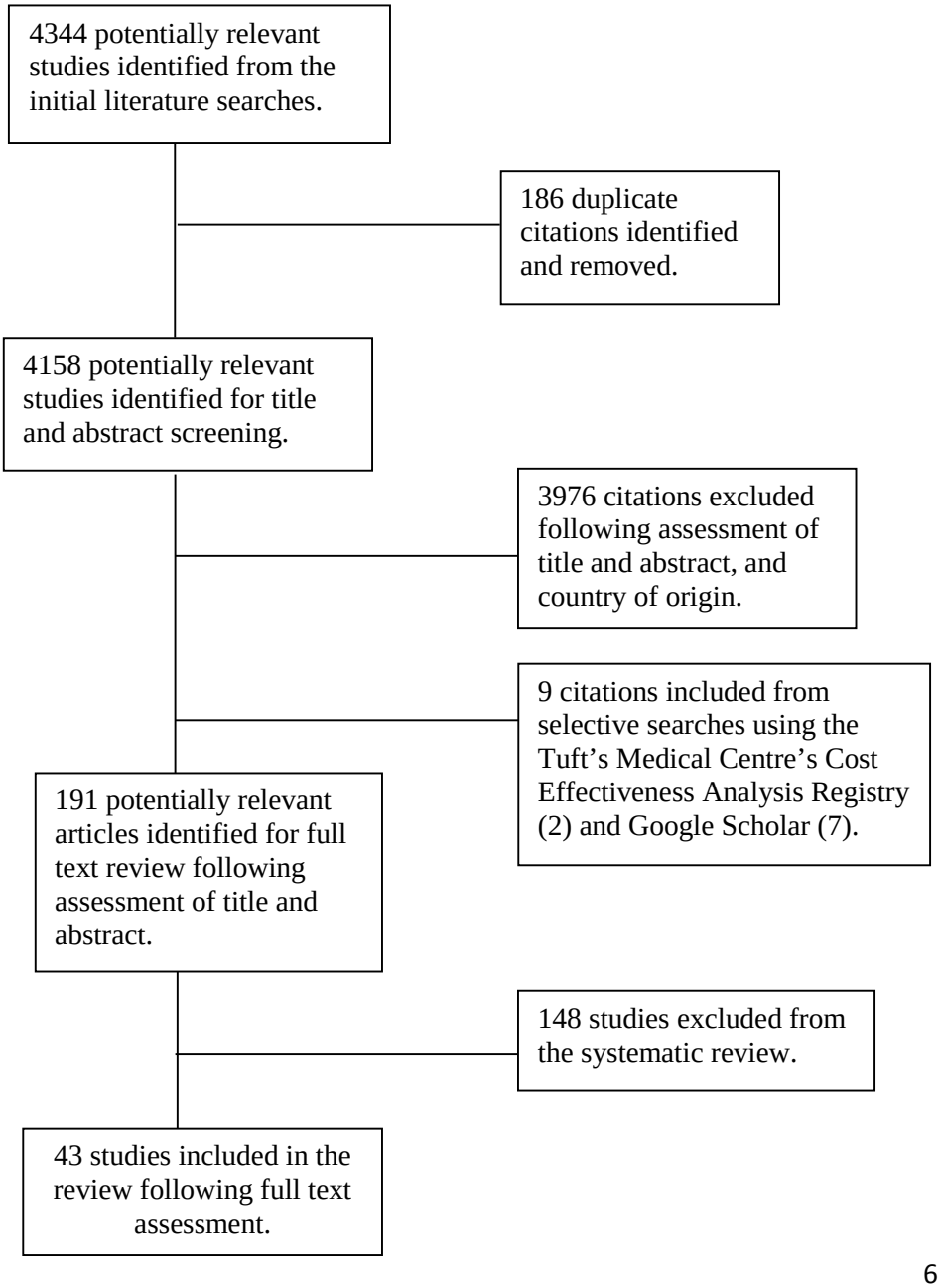
I searched the following electronic databases for relevant studies: the Centre for Review and Dissemination databases (DARE, NHS EED and HTA); EMBASE (The Excerpta Medica Database); MEDLINE using the OVID interface; and Econlit (EBSCO). I used a number of search terms to identify studies providing estimates of the administration costs of biologics. The search terms used and the number of articles retrieved are in Appendix 3A. Additional searches conducted using the Tuft's Medical Centre's Cost Effectiveness Analysis (CEA) Registry and Google Scholar employed terms that included: "time", "motion", "drug", "administration" and "costs". The terms "time" and "motion" were included to identify time-and-motion studies that estimate the direct costs of drug administration by measuring the time it takes to gather materials, prepare medications and administer them to patients. I combined the various search outputs, and titles, abstracts and full texts of the resulting set of articles were assessed. Details of the studies identified and selected are shown below in Figure 3.2.

In selecting studies for full text assessment, I did not set any limits on the year of publication and used the following exclusion criteria when assessing the title, abstract and full text of the articles retrieved. I excluded: non-UK studies; non-English language studies; conference abstracts; studies that do not report on drug administration costs; studies that do not study biologics for human use; and newspaper articles. I excluded studies that did not consider or include drug administration costs in their analysis. I considered that these studies would have underestimated total treatment costs and overestimated cost effectiveness, where it is evaluated. I also excluded studies that claim to account for differences in administration costs but then estimates were not reported, or where administration costs were considered but ignored. For these studies, one could not tell with surety what costs have been included or excluded.

For the purpose of this review, I define biologics using the European Medicines Agency (EMEA) definition stipulated under Directive 2001/83/EC as active substances or medicinal products derived from a biological living system (plant, animal or microbial organism) no matter how 'small' the molecules are. This definition covers not just recombinant proteins, monoclonal antibodies, blood products, immunological products (vaccines, allergens, sera etc.) and advanced technologies (gene and cell therapy products) but also analogous semi-synthetic biologically-active substances or medicinal products of

biological origins that do not belong to any of the main groups mentioned. Because of the broad EMEA definition adopted, technologies evaluated in the set of selected studies are not restricted to commonly referred macromolecular (protein-based) products but also 'biologics' or drugs that might be considered as 'small molecules'. This review included studies that compared a biologic (or a combination of biologic-drug or biologic-device or biologic-drug-device) versus (1) single therapies or combinations of biologics, or (2) small-molecule drugs or drug-device combinations. Also included in the review are studies comparing different drug delivery systems irrespective of whether the active substance is a small-molecule chemically-synthesized drug or a biologic.

Figure 3.2: Study selection



3.3.2 Data extraction

For the selected studies, I extracted variables relating to: study design and methodology; disease condition and patient population; interventions and the treatment alternatives compared; components of drug administration cost (proximal and physical administration); estimates of cost of administration (and the variation around this estimate if reported); and the year to which costs apply. To ensure comparability of estimates of costs of drug administration extracted, I employed the UK NHS Hospital & Community Health Services (HCHS) Index[57], which is a weighted average of the two separate indices the *Pay Cost Index* and the *Health Service Cost Index* to inflate and express all administration cost estimates extracted from the selected studies to year 2010 values.

3.4 Results

3.4.1 Description of studies

I obtained 4344 potentially relevant studies from my searches, of which 186 studies were identified as duplicates (Figure 3.2). From the remaining 4158 studies, 3976 were excluded upon assessment of the titles and abstracts. The main reasons for exclusion were if studies were conference abstracts or newspaper articles, review papers or editorials; and if studies were not written in English or that the country-setting was not the UK. From the 191 potentially relevant studies I identified for full text assessment, 43 studies were included in the literature review and 148 studies were excluded on the basis that they did not provide evidence specific to the administration and use of biologics within the UK NHS.

Of the 43 studies reviewed, 15 evaluated anticancer treatments; nine evaluated vaccination and immunization programmes; eight focused on tumour-necrosis-factor inhibitors for the treatment of psoriatic arthritis, rheumatoid arthritis and Crohn's disease; four focused on prophylactic treatments, (two of which were anticoagulants, one was for prevention of severe respiratory syncytial virus infection and the other study evaluated prophylactic treatment for haemolytic disease); three evaluated insulin analogues; and the remaining studies covered infertility treatments, antiviral treatment of hepatitis C, treatments for Parkinson's disease and for the control of bleeding in people with severe blunt trauma (see Appendix 3B).

3.4.2 Estimates of administration costs for biologics within the UK NHS

From the 43 selected studies, I counted the number of instances 'biologics' were administered by i.v., s.c., i.m., or oral routes. As expected most of the biologics studied were administered intravenously. Appendix 3B shows that administration costs of health technologies given intravenously tend to be higher than those administered subcutaneously; which is higher than those given intramuscularly, which in turn is higher than technologies given orally. For example, the administration cost of enoxaparin given s.c. is £26 whilst the administration costs for standard unfractionated heparin given i.v. is £104[58]. Administration cost of 60mg oral vinorelbine (VNB) is £845, while the administration cost of i.v. VNB (25mg) is £1115[59].

There were differences in the reviewed studies in the components of administration costs that were included (see section 3.4.3) and variations in how studies report drug administration costs. Notably, different measurement scales were used. For instance, some studies report drug administration costs on a per patient basis, while in other studies, costs are reported on a per treatment basis over different time periods; some studies simply report administration costs without specifying the measurement scale.

Taking into account the difficulties of finding a common measure of administration costs across all studies, I computed average estimates of administration costs across a selective set of twenty-four studies that reported estimates on a comparable scale, namely per unit of administration, e.g., per single injection or infusion. Based on the figures extracted from the selected studies, it appears that giving biologics intravenously incurs the highest administration costs while those for subcutaneous and intramuscular routes are of the same order of magnitude (see Table 3.1 below). The average administration costs (in year 2010 values) for 'biologics' administered intravenously is £213, which is almost six to eight times that for biologics given subcutaneously (£35) or intramuscularly (£27). Because these estimates are not derived from statistical analysis of primary data collected from the UK NHS, the figures should be considered as rough approximations of how much it costs to administer biopharmaceuticals via the various routes.

It might be argued that my approximations of administration costs for biologics within the UK NHS were taken from studies comparing different technologies for the treatment of different diseases and clinically indicated for different patient populations. My estimates may therefore hide variation in drug administration costs by disease condition. For any biologic, clinically indicated for two independent disease conditions, one would expect differences in administration costs if the dosage regimens for the treatment of these diseases are different.

However, from the studies reviewed, I did not find any consistent evidence of variation in drug administration costs for a given biologic by disease condition.

To illustrate, infliximab when used in rheumatoid arthritis has administration costs of £148 per infusion in Chen et al.[60] and £175 per infusion in Jobanputra et al.[61]. In Cummins et al.[62], infliximab has administration costs of £128 when used in psoriatic arthritis. Note that while these adjusted figures (in year 2010 values) are quite close to each other, the unadjusted values reported in these studies are the same irrespective of whether infliximab is used for rheumatoid or psoriatic arthritis. Similarly, in Woolacott et al.[63] and Bravo et al.[64], infliximab has administration costs of £307 per infusion when used for psoriatic arthritis but in Dretzke et al.[65], infliximab administration cost is £285 per infusion when used in Crohn's disease. However, the unadjusted values reported in the studies are the same irrespective of whether infliximab is used for psoriatic arthritis or Crohn's disease. On the other hand, annual administration costs for rituximab per patient varies from £369 (corresponding to two infusions) when used in rheumatoid arthritis[66] to £925 (7.4 infusions or cycles) when used in combination with cyclophosphamide, doxorubicin, vincristine and prednisone therapy for diffuse large B-cell lymphoma[67] to £1271 (six cycles) when used in non-Hodgkin's lymphoma[68]. The unadjusted values reported in these studies also vary.

3.4.3 How drug administration costs have been measured

I found that across the 43 studies reviewed there was considerable variation in the components of administration costs specified or included.

Table 3.1: Average costs per injection, infusion or unit administration

Route	Estimate	Comments
Intravenous	£213	This is based on the infusion cost component of administration cost for a combination of 5-fluorouracil and leucovorin (£201)[69]; infliximab infusion cost (£148)[60]; administration cost of docetaxel per visit (£23)[70]; infusion costs of temsirolimus (£203) reported in Hoyle et al.[71]; administration cost of docetaxel (£209)[72]; Cost of hospital stay for chemotherapy delivery (£355) and cost of outpatient attendance for chemotherapy (£292.41)[73]; cost of infliximab infusion (£307 = half rheumatology day case)[74],[75]; administration costs of gemcitabine (£153.50 = cost of medical oncology outpatient visit); and infusion cost component of administration costs of protracted infusion 5-fluorouracil (£235)[76].
Subcutaneous	£35	This is based on the administration cost of enoxaparin (£26)[58]; administration costs of interferon alfa (£28)[77]; injection cost component (£25.33) of the administration cost for gonadotropins[78]; administration cost of peginterferon alfa (£59)[79]; average administration costs of enoxaparin (£13.75)[80]; per weekly administration cost of etanercept (£24.42) derived from the figure (£293) reported in Bravo et al.[75] and Woolacott et al.[54]; and per weekly administration cost (£34) of etanercept and adalimumab (£68) derived from the figure reported in Cummins et al.[62].
Intramuscular	£27	This is based on the administration costs of anti-D immunoglobulin G (£12)[81]; administration costs of hepatitis B vaccine (£20)[82]; administration cost for hepatitis B vaccine (£5.40)[83]; administration costs of influenza vaccine (£7.78)[84]; administration costs of 7-valent pneumococcal conjugate vaccine (£13)[85]; administration costs of vaccines for varicella infection (£13)[86]; administration costs of meningococcal serogroup C conjugate vaccine (£8.80)[87]; administration costs of antenatal anti-D prophylaxis (£4.40)[88]; and administration costs of palivizumab (£161)[89].
Oral	£10	This is based on three studies Cassidy et al.[17], Hoyle et al.[71] and Martin et al.[79] that report zero administration costs; the administration costs for rotavirus vaccines (£6.10)[90] and the unit administration cost for oral chemotherapies (£42)[59].

Le et al.[59], for example, report administration costs of i.v. chemotherapy regimens for non-small cell lung cancer as consisting of cost of day-time hospitalization (£422) and the cost of medical oncology outpatient visit (£143). For oral chemotherapy regimens, administration costs consisted of the cost of pre-therapy counselling with a hospital nurse (£18) and the cost of a general practitioner (GP) to do a local blood test (£24) but zero costs

for physical administration. In Ward et al.[76], the administration costs assigned to i.v. gemcitabine was simply the costs of medical oncology outpatient visits (£153.50). On the other hand, administration costs assigned to 5-fluorouracil given by protracted venous infusion (£1181) consisted of the costs of insertion and removal of a central line (as day case) and the cost of checking and flushing the central line or infusion pump (£235), the costs of hospitalization for drug administration (£711), operations (£209) and concomitant medications (£26.10). In Knight et al.[67], chemotherapy administration costs consisted of pharmacy cost of dispensing and doctor/nurse costs for drug administration. In Sweetenham et al.[68], the administration costs per patient for rituximab consisted of inpatient stay (£423), outpatient visits (£393) and tests (£455). In this study, test costs covered both tests for chemotherapy administration and tests associated with adverse events because of problems of separating the costs of these tests.

In Lloyd et al.[78], administration costs for s.c. human menopausal gonadotropin and s.c. follicule stimulating hormone consisted of staff costs of administering injections (£25.33) and the costs of making a clinic visit for drug administration (£28.20). In Woolacott et al.[74], the s.c. administration costs assigned to etanercept within an initial 3-month period (£293) consists of the first educational visit for etanercept self-injection (the cost of which is equated to one outpatient rheumatology attendance [£131]) and four visits to a staff nurse (each visit costing £40.50) to check on treatment progress. Zero cost is assigned to the physical administration of etanercept. See also Bravo et al.[23]. In Wolowacz et al.[80], administration costs assigned to s.c. low molecular weight heparin (LMWH) is made up of the cost of inpatient visit of £0.85 equivalent to roughly two minutes of nurse time for administering LMWH injections. (This applies to both people eligible for total knee replacement and total hip replacement [THR].) In the case of people undergoing THR, administration costs also included the costs of outpatient visits equal to the cost of district nurse visit per day at £26. For people able to self-administer LMWH s.c. a one-off cost for 30 minutes of nurse time at £11.92 for training in self-administration is incurred.

Differences in how drug administration costs are measured are also evident in Table 3.1, where orally administered medications are either assigned zero or positive costs. Zero administration costs for oral medicines are plausible when one focuses on just the costs of physical administration. That is to say, it costs nothing to swallow a tablet or capsule. If drug administration costs were measured to include all relevant costs of physically administering a drug and the associated proximal costs, one will expect some positive administration costs for oral treatments. For example, in Jit et al.[90], the oral rotavirus vaccines, Rotarix® and

RotaTeq® incur an administration cost equal to 10 minutes of nurse practice time. In Le et al.[59], oral chemotherapies studied have non-zero administration costs covering the costs of outpatient visits and home care; for example, the administration cost of oral VNB is not zero as self-administration of the oral therapy at home requires pre-therapy counselling with a hospital nurse and the cost of a GP to carry out blood tests.

But even if analysts and researchers focus on just physical administration, zero administration costs are not associated with *only* oral treatments. For instance, Morris et al.[91] assign zero administration cost to eptacog alfa (NovoSeven®) although it is administered as an i.v. bolus injection. This (it appears) is because each pack of NovoSeven® is sold as a bundled product consisting of one vial containing the drug, one vial containing sterile water for reconstitution, one sterile vial adapter, a sterile syringe for reconstitution and administration, a sterile infusion set and alcohol swabs. Although one can envisage some staff costs involved in the administration of eptacog alfa, it easy to see why zero administration costs is not too unreasonable.

Indeed, a look at the spread of administration costs for each injectable route in Table 3.1 shows some overlap although average estimates differ. One explanation is that there is legitimate variation in the methods of drug administration, and this is reflected in the different studies. This seems unlikely given the focus on studies undertaken in a single country. An alternative explanation is that studies are not providing a complete or consistent account of the costs of physically administering biologics and the proximal costs involved. There are 'hidden' costs of drug administration that some studies fail to consider. Given the observed variation in how different studies define the components of administration costs, the reported average estimates are necessarily imprecise (perhaps conservative) approximations.

3.5 Discussion

Two things can be gleaned from this review. One, administration costs are an important component of assessing the incremental cost-effectiveness of a given medical technology relative to its comparators. Two, drug administration costs consist of the costs of physical administration plus proximal costs incurred before and after physical administration of a medicine. I found that there is variation in the range of administration costs included. This review shows that the definition of administration costs for biologics is critical. Not only do administration costs vary by route of administration but, for any given route of drug administration, there is variation in costs as a function of the method of measurement.

Differences in how studies define administration costs seems contrary to recommendations of various methodological guidelines and quality checklists[40] that *all* relevant costs, beyond simple acquisition costs, should be identified, quantified and reported. A review of pharmacoeconomic evaluations conducted in Ireland and the UK by Hughes et al.[93] showed that reporting and conduct of cost estimation is in need of improvement. The authors suggest a checklist of items that, among other things, requires both the route of administration and costs relating to drug administration are reported. This checklist, however, separates drug administration costs from "pharmacy charges" covering professional and dispensing fees; "hospital pharmacy costs" covering the costs of reconstituting parenteral preparations etc.; and "therapeutic drug monitoring costs" covering, for e.g., biochemical assays or measurements of plasma concentrations and pharmacodynamic responses. If researchers are to use the framework presented here as part of or in addition to their guidelines and quality checklist(s), these items will automatically be incorporated into estimates of drug administration costs. This will ensure consistency in what cost components are included or excluded. It must be emphasized that, although the focus was on UK studies, the framework depicted in Figure 3.1 is 'global' and applicable to non-UK settings.

Based on this review, administration costs of biologics given intravenously appears to be, on average, six or eight times those given subcutaneously or intramuscularly; and biologics administered orally tend to be assigned zero administration costs. Admittedly, these average estimates suffer from being taken from a selected sample of studies that offer a comparable scale of measurement; and of course, the literature data are not necessarily correct given variations in how the components of drug administration costs are defined across studies. These average estimates, nevertheless, provide a reasonable guide as to the costs of administering biologic products within the UK NHS. Yet even for i.v. biologics, near zero administration costs are possible if manufacturers sell biologics as bundled products with some of the equipment and consumables employed in administering these products. There are several rationales for product bundling[94] but my findings suggest there is an additional advantage from selling biologic products bundled with equipment and consumables employed in drug administration. This stems from lower administration costs; to be precise, lower proximal costs of gathering materials and preparing medicines for administration to patients. Granted, one would expect that, in most cases, the costs of equipment and consumables will be reflected in the price of the bundled product even if bundling was intended to conceal price discounts on a drug product that could be sold separately.

However, my rough estimates of drug administration costs do not consider increments or decrements in (administration) costs that might be associated with compliance issues and medication errors. Webster et al.'s[95] comparison of a Mark II (pre-filled syringe) system with conventional methods of administering various drugs used in cardiac anaesthesia indicates that although use of new safety-oriented administration system saved preparation time before and after anaesthesia, it increased overall costs per anaesthetic administered (median [range] €178 [102–428] vs. €155 [111–390], p = 0.041). The savings in administration costs was lower than the purchase cost of the Mark II system. However, this cost increment is likely to be offset by potential reductions in iatrogenic events via reducing administration errors, i.e., avoiding cuts in glass ampoules and needle-stick injuries; violation of sterile techniques during draw up of medicines; injection of minute glass shards and mislabelling of syringes under the pressure of an operating theatre.

Similarly, Kruse et al.[96] argue that although a move away from i.v. single-agent therapies for breast cancer to oral formulations may save on administration costs, these savings may be offset by the added costs associated with oral agents such as noncompliance or treatment of gastrointestinal and other side effects. Also Harley et al.[97] found that, in people with rheumatoid arthritis, compliance with, at least, 80% of doses administered was lower for etanercept (odds ratio of 0.462, 95% CI: 0.290–0.736) and for methotrexate (odds ratio of 0.385, 95% CI: 0.245–0.604) when compared to infliximab. This possibly reflects differences in the route and frequency of administration: infliximab is given intravenously every 8 weeks, etanercept is given subcutaneously twice a week and patients anxious about self-administering s.c. therapies tend to miss injections and prefer i.v. infusions. The authors further argue that although other studies report higher administration costs for infliximab, their study show infliximab was associated with lower facility costs (i.e., physician costs and other fees related to drug administration) compared to etanercept. This may be due to better dosage compliance with infliximab (besides problems of small sample sizes, inferring drug usage rates from insurance claims data, and of course, differences in how studies measure administration costs).

So what are the implications for (research in) biopharmaceutical manufacturing? Consider a pharmaceutical manufacturer *X* who has developed a biologic product *Y* and intends to launch this product on the UK NHS market. Following the work of Vernon et al.[4], there are substantial private and societal gains in the short or long-run from investments to lower the cost of goods (COGs) for product *Y*. In the short-run, lower manufacturing costs, may lead to lower prices and increased consumption, especially when

faced with competing products. However, even if manufacturer *X* doesn't lower prices but keeps the cost savings, this might still improve social welfare in the long-run as it is known that increments (decrements) in gross price-cost margins are associated with higher (lower) intensities of R&D[18]. That is, part of the accumulated savings or 'profits' from reducing COGs over time will be used to finance future R&D and product innovations. Indeed, Vernon et al.[4] report that the societal gains from keeping the cost savings could be as much as twelve times that of the gains from passing cost savings on to consumers in the form of lower prices. However, that argument assumes existing demands for product *Y* expressed by a healthcare payer who is concerned with only net clinical benefits offered by product *Y* (efficacy minus safety) and its acquisition costs; and not the total delivery costs to the healthcare system.

Currently the UK NHS relies on NICE to make technology adoption decisions on its behalf, on the basis of cost-effectiveness evidence, besides other factors. Following Dakin et al.[98], the outcomes of NICE's decision-making can be categorized as "Yes" (i.e., recommended for use within the EMEA approved marketing authorization), "No" (i.e., not recommended and in which case, demand effectively falls to zero) and "Yes, but" (i.e., recommended for restricted use – in which case demand will be less than anticipated prior to market launch). In the worst-case scenario of zero demand, there will be a costly loss of on-patent sales revenue; and no societal gains in the short- or long-run from process innovations to reduce COGs. Things get better with a "restricted" demand for product *Y* but this still falls short of what is attainable in the best-case scenario. But the chosen formulation for product *Y* and the associated administration costs, if the full range of costs is measured, may be the key determinant of whether product *Y* is considered cost-effective or not; the level of demand for product *Y* and the flow of clinical benefits to patients (see also Cheng et al.[99]). This means that besides concentrating on selecting biologic drug candidates on the basis of net clinical benefits, "getting a product out", and ensuring the COGs for biologic products is below defined (company-specific) thresholds or rules-of-thumb, manufacturer *X* should pay equal attention to issues around formulation, drug administration costs and total healthcare delivery costs of product *Y* as early as possible over the R&D timescale.

I do not know for sure whether manufacturers of biologics routinely consider administration costs and total delivery costs that will be borne by healthcare payers when making manufacturing and R&D decisions. It is possible that costs to healthcare payers are seldom considered by biologic companies in R&D and manufacturing decisions (or perhaps do so only when faced with therapeutic competitors); and this might reflect conflicting

internal priorities or organizational disconnection between companies' manufacturing (process development) units and marketing departments. I will argue that manufacturers of biologic medicines should consider formulation issues, drug administration costs and total healthcare delivery costs as early as possible in their manufacturing and R&D decisions.

This study, however, suffers from two limitations. First, this chapter focused on only the cost of administering biologic medicines within the UK NHS. This means my findings may not be applicable to other country contexts or to small-molecule chemically-synthesized medicines. That said the framework of drug administration is still useful for consistent evaluations of drug administration costs irrespective of the country context or whether the focus is on small-molecule chemically-synthesized medicines. A de novo analysis is ultimately needed to assess the magnitude and importance of drug administration costs within the UK NHS, with researchers using Figure 3.1 as a tool to ensure inclusion of all relevant costs of physically administering biologics and the associated proximal costs. Second, I did not assess the studies included in the review for their methodological quality, for sources of systematic error or bias (internal validity), or their external validity. Third, my searches, using the same inclusion criteria, retrieved HTA reports but not NICE technology-appraisal submissions which may have additional evidence on drug administration costs. The reason why NICE submissions were not part of my review is because they are not peer-reviewed or their titles and abstracts did not match my combinations of search terms. These limitations provide yet another reason why the estimates of administration costs for biologics should be considered rough approximations, and why a de novo analysis of administration costs for biologics within the UK NHS is worth pursuing.

Key points

- Drug administration costs include the costs of physically administering a drug, plus proximal costs incurred before or after physical administration.
- Drug administration costs vary by route of administration. On average, administration costs are higher for biologics given intravenously when compared to those given by intramuscular or subcutaneous routes.
- There is variation in the way administration costs have been measured in research studies.
- Formulation issues, drug administration and healthcare delivery costs should be considered as early as possible when making manufacturing and R&D decisions.

3.6 Conclusions

The key points from this literature review are listed in the textbox above. Drug administration costs vary by route of administration with average estimates for intravenous routes higher than that for subcutaneous administration, which is higher than that for intramuscular administration, which in turn is higher than that for oral routes. These findings are, however, tied to the definition of the components of administration costs used in the studies reviewed. Within each route of administration, there is variation in costs as a function of the method of measurement. This matters as estimates of administration costs will differ across studies and according to what cost components are included or excluded. Also, my results may need to be modified in lieu of issues around medication errors and compliance but this study has provided evidence as to how administration costs of biologics can be lowered by selecting the most appropriate dosage form and regimen, and route of administration. In some cases, it is possible to lower administration costs via bundling biologic products with equipment and consumables used in drug administration.

This chapter argues that it is not enough to focus on getting a commercially-viable product out and the market price 'right', especially when faced with healthcare payers who rely on the outputs of economic evaluation to guide their technology adoption decisions. Besides, process innovation efforts to reduce the cost of goods for biologic products, manufacturers should consider (in the early phases of drug development) formulation issues and the impact on administration costs and total costs of healthcare delivery.

4 Development of an Administration Cost Algorithm

4.1 Introduction

The adoption and utilization of beneficial medical technologies including biologic drugs has, in recent times, been subject to a number of regulatory, marketing, reimbursement (coverage) and other demand-side hurdles including the so-called risk-sharing arrangements. The demand-side hurdles, in particular, have evolved out of increasing concerns by healthcare payers about the high acquisition costs and budgetary impacts of these medical technologies. In response to these concerns, healthcare payers, acting as responsible agents on the behalf of their patient populations, have turned to the use of CEA, budget impact analysis or HTA to estimate cost-effectiveness and affordability prior to the deciding whether to adopt or reimburse the usage of biologics within their original or restricted marketing authorization or not. This typically involves comparing, over a specified time period, total disease management or healthcare delivery costs (i.e., the sum of drug acquisition costs plus administration and future healthcare-related costs) associated with a given technology and the total health benefits expected. Holding all else constant, the estimated cost-effectiveness of a biologic drug may be dictated by how much of healthcare resources are spent on drug administration, and whether trade-offs exist between drug acquisition costs and administration costs.

On the supply-side, it has been observed that biological drug candidates are developed with a skewed focus on clinical efficacy and safety to the neglect of issues related to the ease of manufacturing, affordability and cost-effectiveness of these therapies to healthcare payers. This often leads to unnecessary waste and excessive reworking of manufactured products, and a growing concern over the failures and struggles manufacturers face in passing through what is becoming an increasingly complex set of regulatory and demand-restricting hurdles. That latter is known to be associated with significant delays in market launch, in addition to the time and revenue lost in price negotiations[111]. But besides worrying about the ease of manufacturing, it is also useful for manufacturers to, at least, consider prior to market launch, the administration costs (and total healthcare delivery costs) associated with the different ways they choose to manufacture and formulate their products. In that case, manufacturers' evaluation of drug administration costs should be done in the same manner as it will be conducted by healthcare payers. The argument here is not that drug manufacturers pay little attention to routes of drug administration; but that the contribution of administration costs to total healthcare delivery costs should not be trivialized or ignored. Differences in drug acquisition costs could be eroded by differences in administration costs, leaving healthcare delivery costs unchanged.

The aim of this chapter is to conduct to a de novo analysis of administration costs of biologic drugs, to identify the factors that affect variation in these costs, and argue why such evaluations are an important step in the conduct of biopharmaceutical manufacturing prior to market launch. That is, I explore why pharmaceutical manufacturers should consider the link between administration costs (and how this is influenced by formulation and manufacturing), total healthcare delivery costs and value-for-money when making their go-no-go R&D decisions.

This research stream is motivated by two key points. First, in a systematic review of the economic value of reducing medication dosing frequency using drug delivery systems, Cheng et al.[98] found that in most cases drug products with less-frequent dosing schedules tend to be cost effective when compared to conventional (standard) formulations containing the same active moiety – although these 'advanced' or 'improved' delivery systems may be expensive to make.

Second, the systematic review in Chapter 3 of this thesis identified possible trade-offs between acquisition and administration costs: a drug that appears cheap to buy may (in the long-run) have higher total healthcare delivery costs as more NHS resources are spent on drug administration. The budgetary impact of a biologic with high acquisition costs (because of complex and costly manufacturing, for example) but relatively low administration costs could be the same as or lower than that of a less expensive biologic with higher administration costs. Recall this systematic review also found that there are inconsistencies in how studies define administration costs and consequently, differences in the type of costs included or excluded from estimates of drug administration costs. See also Tetteh and Morris[112]. These differences in what cost items are included or excluded from estimates of drug administration costs means some of the reported differentials in administration costs for biologic drugs may not be real and cannot be used unreservedly in economic analyses. Once differences in cost estimates simply reflect differences in methods of measurement, one cannot tell for sure whether trade-offs exist between administration costs and drug acquisition costs or not; and to what extent administration costs could influence conclusions reached about the cost-effectiveness of medical technologies.

Taking into account these points, I evaluated variations in the administration costs for a sample of eighteen biologic drugs listed for use in the UK NHS, taking care to ensure consistent inclusion or exclusion of all relevant costs related to drug administration. I do this to ensure very little variation in drug administration costs can be attributed to differences in the method of measurement.

I then developed an administration-cost algorithm to help manufacturers predict, prior to market launch, the administration costs associated with their formulation choice for each biologic drug candidate in their portfolio. I believe this, together with manufacturer's expectations of product prices, should help them consider the possible trade-offs between drug

acquisition and administration costs; and generate credible estimates of total healthcare delivery costs of their drug products and the likelihood that these products will receive favourable recommendations from healthcare payers. The algorithm allows us to highlight cases where (1) administration costs for a given biologic medicine varies across different clinical indications and patient populations, (2) administration costs for a given biologic medicine differs by the route of administration, (3) differences in drug administration costs for different dosing regimens of a given biologic medicine, and (4) administration costs can be lowered via bundling medicinal products with some of the equipment and consumables used in administering drugs.

The chapter is structured as follows. I first describe in section 4.2 the methodological approach taken, the underlying assumptions made in my analyses, and the data sources used. This is followed by sections 4.3 and 4.4 with results and discussions. Section 4.5 completes this chapter with my conclusions.

4.2. Data and methods

As indicated above, this research aims to evaluate the administration costs of a sample of biologics to identify factors that affect the magnitude of (differences) in drug administration costs. To avoid overcomplicating the analysis, I assume *clinical outcome neutrality* in that for any comparison of different modes of administering the same biologic drug, there are no differences in net health benefits (i.e., efficacy minus safety concerns) or that differences in net health benefits have no bearing on the magnitude or variation in administration costs. For example, differences in the incidence and severity of adverse events between two or more formulations of a given biologic drug will have no bearing as to how much is spent on drug administration costs. I also ignore other costs associated with disease management.

4.2.1 Identifying and measuring administration costs

From an economic perspective, costs measured should reflect the opportunity costs of NHS resources deployed in administering biologic drugs that could otherwise have been used elsewhere had the drug in question not been administered. An accurate measurement of drug administration costs thus requires identifying all resources that will be expended or the ‘cost centres’ where resources will be consumed and costs incurred[113]. To identify the ‘cost centres’ related to the administration of biologic drugs (from a healthcare payer perspective), I employed the framework of drug administration described in Chapter 3, page 59 (see also Tetteh and Morris[112]) that makes a distinction between the proximal costs of drug administration

and the costs of physical administration. In that framework, proximal administration costs (Pc) refer to costs incurred before or after physical administration of the drug into the body whilst physical administration costs (PAc) refer to the costs of physically introducing the drug into a patient via one of the established routes for administration. Each component labelled in that framework constitutes a (micro-level) cost centre where resources are consumed and costs incurred.

I used this framework to ensure consistency in what type of administration costs are included or excluded in the analysis. Using a common yardstick, I believe, should support (1) complete or near complete accounting of the opportunity costs associated with drug administration, and (2) "apples to apples" comparison of the administration costs of biologic drugs such that very little variation in administration costs can be attributed to the method of measurement. For the same reasons, I defined a common time frame over which drug administration costs will be estimated. I chose to evaluate annual costs of drug administration costs as this fitted well with the dosing regimens of all products in my sample. For simplicity, the analysis based on a single patient that successfully completes a single full treatment course over a 12 month period. It might be argued that this will introduce bias against biologics indicated for an acute illness with typically 'short' treatment episodes. However, extending the time frame beyond one-year period will actually amplify the cost differences between acute and chronic biologics whilst a 6-month period will not fit with the dosing regimen of some of the products in my sample. What is more, I do not consider repeat treatment episodes over the one-year period. Hence, my estimates of annual drug administration costs should not be biased against biologics indicated for acute illnesses.

4.2.2 Product sample and dosing regimens modelled

My analysis makes use of an unbalanced sample of 18 therapeutically-active biologics; of which eight are administered intravenously, eight given by subcutaneous delivery and two given intramuscularly. Within this sample, fifteen of the products are humanized monoclonal antibodies (mAb) with the remainder comprising of one fragmented monoclonal antibody (fAb), a fusion protein and an interferon protein. The characteristics of this product sample are presented in Table 4.1 below. This sample follows a list prepared under the research flagships of the EPSRC Centre for Innovative Manufacturing in Emergent Macromolecular Therapies. I make no argument that this sample is representative of all existing or emergent biologics or macromolecular therapies. It is possible that this selective list excludes some important biologic therapies. Although there are a host of orally-administered semi-synthetic biologically-derived

medicines, none of the biologic therapies in Table 4.1, at the time of conducting this research, were given orally. The general argument is that biologic drugs have poor bioavailability when swallowed. That said, there is a lot of research underway to develop biologics for oral administration – and the cost algorithm presented in this chapter could be easily updated in future research with an expanded sample of biologics including those given orally.

Table 4.1: Characteristics of product sample

Product name	Product type	Clinical indication considered	Disease type
Biologics given intravenously			
Basiliximab†, Simulect®	mAb	Prophylaxis of acute organ rejection in allogeneic renal transplantation.	Acute
Bevacizumab, Avastin®	mAb	First-line treatment of adult patients with non-small cell lung cancer.	Chronic
Cetuximab, Erbitux®	mAb	Treatment of epidermal-growth-factor-receptor (EGFR)-expressing, Kirsten Rat Sarcoma-2 (KRAS) wild-type metastatic colorectal cancer.	Chronic
Infliximab, Remicade®	mAb	Treatment of active rheumatoid arthritis that is unresponsive to disease modifying anti-rheumatic drugs (DMARDs) or those with severe active disease not previously treated with methotrexate (MTX) or DMARDs.	Chronic
Oftamumab†, Arzerra®	mAb	Treatment of chronic lymphocytic leukaemia that is refractory to fludarabine and alemtuzumab.	Chronic
Panitumumab, Vectibix®	mAb	Treatment of wild-type KRAS metastatic colorectal cancer.	Chronic
Tocilizumab, Actemra®	mAb	Treatment of moderate-to-severe active rheumatoid arthritis.	Chronic
Trastuzumab, Herceptin®	mAb	Treatment of advanced and metastatic breast cancer.	Chronic
Biologics given subcutaneously			
Adalimumab†, Humira®	mAb	Treatment of moderate-to-severely active Crohn's disease in children.	Chronic
Canakinumab†, Ilaris®	mAb	Treatment of cryopyrin-associated periodic syndromes (CAPS) in adults.	Chronic
Certolizumab pegol†, Cimzia®	fAb	Treatment of moderate-to-severe active rheumatoid arthritis in adults when response to DMARDs including MTX has been inadequate.	Chronic
Denosumab†, Prolia®	mAb	Treatment of osteoporosis in post-menopausal women at increased risk of fractures.	Chronic
Etanercept†, Enbrel®	Fusion protein	Treatment of active and progressive psoriatic arthritis that has not responded adequately to DMARDs.	Chronic
Golimumab†, Simponi®	mAb	Treatment of moderate-to-severe active ankylosing spondylitis.	Chronic
Omalizumab†, Xolair®	mAb	Management of immunoglobulin-E mediated asthma.	Chronic
Ustekinumab†, Stelara®	mAb	Treatment of moderate-to-severe plaque psoriasis.	Chronic
Biologics given intramuscularly			
Palivizumab†, Synagis®	mAb	Prevention of serious lower respiratory tract disease caused by respiratory syncytial virus (RSV) in children at high risk of RSV disease.	Acute
Interferon beta-1a†, Avonex®	Interferon protein	Treatment of relapsing multiple sclerosis in adult patients, i.e., two or more acute exacerbations in the previous three years without evidence of progressive disease.	Chronic

Notes: † refers to biologics that are sold bundled with some of the equipment and consumables used in drug administration.

To estimate the costs of administering the biologic drugs in my sample, some idea or knowledge of the dosing regimen for patients considered eligible to receive a given biologic drug is needed. I used the dosing regimen indicated by the marketing authorisation for a given biologic drug, gathering this information from the posology described in the drug's package inserts, the summary of product characteristics (S*m*PC) posted on the European Medicines Agency (EMA) website; the British National Formulary (BNF) or the electronic Medicines Compendium (eMC). Obviously, within UK NHS settings, the prescribed pathway suggested by the regulatory license or marketing authorization may not necessarily coincide with actual clinical practice – bearing in mind possible gaps between recommendations in HTA guidance and clinical guidelines and implementation of these recommendations in routine practice; as well as practice-specific watch-and-wait treatment strategies.

To avoid the complexity introduced by what happens in routine clinical practice, I simply modelled the drug administration instructions given in the products' package inserts or the S*m*PC. For this, I assumed continuous dosing of a given biologic for the whole year unless the marketing authorization or S*m*PC provides clear indications as to the maximum number of doses or recommended duration of treatment. This is because I found it hard to make any unquestionable assumption about the proportion of treatment-responders and non-responders.

4.2.3 Analysis

I first conducted a deterministic analysis with the estimated costs of drug administration disaggregated into proximal costs and the cost of physically administering the drug. I used the framework described in Chapter 3, page 59, as a guide to selecting which cost item to include as long as there was publicly-available data for that cost item. My analysis was done using an economic model built in Microsoft Excel with data inputs from the BNF and eMC, the NHS Reference Costs 2011-2012, the NHS electronic Drug Tariff and the 2012 edition of the PSSRU (Personal Social Services Research Unit) Costs for Health and Social Care, and from published and grey literature.

One issue with such deterministic analysis is the well-known fact that there is: (1) uncertainty in the incidence and severity of illness and for that matter, demands for health intervention using biologic drugs; plus (2) uncertainty with regards to treatment outcomes, which fuels future demands for healthcare intervention; for example, dose reduction or escalation; modifications to dosing regimens and treatment protocols and/or deployment of alternative (salvage) interventions in complementary or substitutive ways[114],[115]. For this reason, the quantity and costs of NHS resources expended on the administration of biologic

drugs cannot be described by fixed values – considering also deviations of what happens in routine clinical practice from the EMA-approved posology.

I resolved this issue by introducing parameter uncertainty via fitting a gamma distribution to the deterministic estimates of proximal costs (Pc) as well as the costs of physical administration (PAc), and running 1000 Monte Carlo simulations[9] for each product. Note that the choice of a gamma distribution is not arbitrary but reflects the observation that healthcare resource use and costs are skewed with non-negative values ranging from zero to positive infinity. As argued by Nixon and Thompson[116], skewed parametric (gamma, log-logistic, lognormal) distributions fit medical cost data better than a normal distribution and should in principle be preferred for estimation. What is not clear, however, is which skewed parametric distribution is best. Here I chose a two-parameter gamma distribution, and given the absence of real-life data that reflects the uncertainty of healthcare demands requiring intervention with biologics, I adopted the simplest assumption that the mean and standard error (SE) for administration costs for any given biologic drug in my sample is the same (see Briggs et al.[117]). I do not expect this to introduce any systematic bias as the assumption will apply to all products within the sample.

From the synthetic dataset generated, I computed summary statistics including 95% confidence intervals for mean expected administration cost for each biologic drug. I also identified what proportion of total administration costs are due to proximal costs or physical administration costs for my sample of biologic drugs (categorized according to their respective routes of administration) by graphing a log-log plot of Pc versus PAc. I took logs of proximal and physical administration costs because of the skewness of a gamma-distributed cost data and to narrow down the range of (large) values. And estimated the proportion of simulations where the ratio of proximal costs to physical administration costs (PAc/Pc) is greater than or equal to 1. The essence of this exercise is to identify the source of cost savings from changes in drug formulation and manufacturing.

4.2.4 Identifying the algorithm

Recall that one of my research objectives is to develop an algorithm that will allow manufacturers to predict how administration costs change with how they choose to manufacture and formulate a given biologic drug candidate. On *a priori* grounds, I defined this algorithm as a

[9] The choice of 1000 simulations is not arbitrary. A simulation convergence test, which is not reported in the paper, suggests a virtually flat administration-cost curve with the number of simulation trials ranging from 1000 to 200,000. I do this by running each 1000 simulation 200 times.

population regression function linking drug administration costs (ADMINCOST) and a number of independent explanatory variables (X). I considered my simulated (synthesized) dataset as one of a number of finite-sized datasets (samples) randomly drawn from this population distribution of biologic administration costs, which can be described by $(X_1, X_2, \dots, X_J; ADMINCOST)j$ and the data-generating function: $ADMINCOST = g(X_1, X_2, \dots, X_J;\ \beta_0, \beta_1, \dots, \beta_k)j$, where j = 1, 2,…, J is an index for each biologic drug, the β's are parameters to be estimated and k is an index for each explanatory variable.

The explanatory variables I chose are those that I believe (bio)pharmaceutical manufacturers will have some information on, or an idea of, as they work on a number of promising biologic drug candidates, and decide which candidates to take forward to the next stage of process R&D or product development, and which ones to reserve as contingency or backup options. Given the skewness of the simulated data for ADMINCOST (on the raw scale), I estimated the following (linear) sample regression function with a log-transformed dependent variable:

$$\begin{aligned} ln(ADMINCOST) &= \alpha + \beta_0 ROUTEADMIN1 + \beta_1 ROUTEADMIN2 + \beta_2 DOSFREQ \\ &+ \beta_3 PRODUCTBUND + \beta_4 INDICATN + \beta_5 DOSFREQ^2 \\ &+ \beta_6 DOSFREQ.INDICATN + \epsilon \end{aligned} \tag{9}$$

The intercept is α and the error term (ϵ) represents any unexplained variation in administration costs.

ROUTEADMIN is a three-group categorical variable indicating whether a given biologic drug is administered intravenously, subcutaneously or intramuscularly. This, in the sample regression function, is implemented as two separate dummy (zero-one) variables: ROUTEADMIN1 (which takes the value of one if a product is given subcutaneously and zero otherwise) and ROUTEADMIN2 (which takes on the value of one if a product is administered intramuscularly and zero otherwise). Intravenous administration is therefore the baseline, benchmark or reference category for the variable ROUTEADMIN. I do this because for most biologics, intravenous infusion is the default (conventional) drug delivery or formulation choice given the fragility and poor oral bioavailability of macromolecular proteins. The second variable DOSFREQ refers to the frequency (intensity) of dosing, which I define as the number of 'unit administrations' in a given year. PRODUCTBUND is also a zero-one dummy variable indicating whether a given biologic product is sold together with some of the equipment and

consumables used in drug administration (1) or not (0). Similarly, INDICATN is a zero-one dummy variable indicating whether a biologic drug is for the management of an acute illness (0) or for a chronic illness (1). The interaction term DOSFREQ.INDICATN is intended to capture the notion that treatments for acute illnesses tend to have less 'complex' dosing regimens compared to those for chronic illnesses; and the quadratic term $DOSFREQ^2$ is intended to determine whether the marginal effects of dosing frequency are increasing or diminishing as the frequency (intensity) of dosing increases.

I estimated the four nested models using *linear* ordinary least squares (OLS) regression. I labelled the four nested models as A, B, C and D – all of which rely on different combinations of the explanatory variables defined above. I designated C as the full unrestricted model as it includes all the explanatory variables above. Model A only considers variables for subcutaneous and intramuscular routes and dosing frequency. Model B adds variables for bundled products and acute/chronic illness to the variables in model A whilst model D includes all the variables in model C with the exception of the variables for subcutaneous and intramuscular routes. Model D is designed to test the relationship between the variable for acute/chronic illness and the variables for route of administration. I initially performed the log-OLS regressions in Excel using the LINEST function. The main advantage of this approach is that, because one is working from the same software platform, the analysis captures (minor) differences in expected mean ADMINCOST for every 1000 simulations. That is, there is a direct dynamic connection between the simulated dataset, the estimated parameters and predicted values from the log-OLS regression.

4.2.5 Robustness and validation

A well-known problem with OLS regression using a log-transformed dependent variable is that retransformation of estimates of the logged cost of drug administration, $\ln\widehat{ADMINCOST}$, to the raw untransformed scale gives the geometric mean for $\widehat{ADMINCOST}$ (which is often close to the median) rather than the arithmetic mean, the parameter of interest. (The hat on ADMINCOST indicates an estimated or predicted value of drug administration costs.) The problem is resolved by using what is called a smearing factor to minimize the prediction error. The name is derived from the fact that the factor distributes (smears) the 'excess' or prediction error in one observation to other observations proportionally when adjusting unlogged median estimates to unlogged mean estimates. Generally, this takes the form: $\widehat{ADMINCOST} = \exp(X\hat{\beta}_k)\bullet\hat{\Phi}$, where $\hat{\Phi}$ is the smearing factor. I considered three ways of deriving less-biased smearing estimates on the raw untransformed scale.

The first method yields what is called 'normal theory estimates' that are derived under the assumption that the error term for $\ln\widehat{ADMINCOST}$ is normally distributed, in which case the smearing estimate of ADMINCOST = $\exp(X\hat{\beta}_k + 0.5\hat{\sigma}^2)$ where $\widehat{\sigma^2}$ is the square of SE($\ln\widehat{ADMINCOST}$) and X refers to the value(s) for the explanatory variable(s). The smearing factor $\hat{\Phi} = \exp(0.5\hat{\sigma}^2)$. If errors are not normally distributed but are homoscedastic (i.e., have constant variance), then the second method, which uses non-parametric (sub-group specific) smearing factors can be used to minimize prediction errors. The non-parametric smearing factor is given by: $\hat{\Phi} = N^{-1}\sum_{i=1}^{N}\exp(\widehat{\epsilon_i})$, where N is the number of observations and $\widehat{\epsilon_i}$ is the log-scale residuals from the regression. When the errors are heteroskedastic (i.e., non-constant variance), or they depend on the explanatory variables, it is "better" to use a subgroup-specific smearing factor for each intravenous, subcutaneous or intramuscular product category[118],[119],[120]. The third method uses a regression-through-the-origin approach suggested by Wooldridge[112]. This is as follows: obtain for each observation the naïve estimates $\widehat{m_i}$ = $\exp(ln\widehat{ADMINCOST})$; then perform a regression of ADMINCOST on $\widehat{m_i}$ through the origin and obtain the only coefficient $\widehat{\alpha_0}$ as the smearing factor.

Irrespective of the smearing factor used, its value in minimizing prediction error depends crucially on the presence and nature of heteroskedasticity in the log scale residuals. For example, the subgroup smearing factors assume that log scale heteroskedasticity varies across the mutually-exclusive subgroups specified. If, however, heteroskedasticity varies according to one or more (continuous, discrete or interacted) explanatory variables in the OLS regression, then we will have smearing estimates of ADMINCOST that are still biased (prediction losses remain). It is possible to run an auxiliary regression of the heteroskedastic variance as a function of one or more of the explanatory variables, i.e., $\hat{\Phi}$ =$\exp(\widehat{\epsilon_i}) = \rho(X)$; but then this depends on how much of the heteroskedastic variance is explained by the chosen set of explanatory variables. Generally, this auxiliary-regression approach is thought to be cumbersome and there is no simple fix if the form of heteroskedasticity is unknown.

An alternative estimator, however, exists in the form of generalized linear modelling (GLM) to overcome the retransformation problem. GLM does this by directly and independently specifying: (1) a link function between the raw scale ADMINCOST and the linear index ($X\hat{\beta}_k$), and (2) a family of parametric distributions to reflect any heteroskedastic relationship between the raw scale error variance and $\widehat{ADMINCOST}$; i.e., var($ADMINCOST$) $\cong$ $\varphi.[\widehat{ADMINCOST}]^{\delta}$, where δ is the over-dispersion parameter and φ is a constant[120],[121],[123]. Independent specification of the link function (i.e., the scale of estimation) and family distributions (i.e., the variance function) under GLM allows $\widehat{ADMINCOST}$ to be estimated directly (or from the

exponent of ln$\widehat{ADMINCOST}$) without the need for smearing factors. In most applications, three types of link functions are specified: "identity", "log" and "power". Here an identity link specifies the following relationship: $\widehat{ADMINCOST} = X\beta_k$; the log link takes the form $\widehat{ADMINCOST} = \exp(X\beta_k)$ whereas the power link takes the form $\widehat{ADMINCOST} = (X\beta_k)^{\tau}$ where τ is the power specified. The family distributions commonly investigated or used to model heteroskedasticity are Gaussian if the parameter $\delta = 0$; Poisson if $\delta = 1$, Gamma or heteroskedastic normal if $\delta = 2$ and inverse Gaussian if $\delta = 3$. The appropriate family distribution is often identified using the so-called modified Park test, that involves a log-gamma GLM regression of var($ADMINCOST$) on ln$\widehat{ADMINCOST}$. The coefficient on ln$\widehat{ADMINCOST}$approximates $\hat{\delta}$.

The main drawback of these estimation approaches is one loses the dynamic link with the simulated data in Excel – having exported the average expected ADMINCOST over 1000 simulations for each product into STATA. I did not expect this to really make a 'big' difference as we know from a simulation convergence test (see footnote 9 on page 81) that ADMINCOST curve is virtually flat over two hundred 1,000 simulations.

Notwithstanding, GLM is known to suffer prediction losses if one has heavy tailed data (kurtosis) even after log retransformation of the dependent variable. GLM prediction losses (relative to the log-OLS estimator) increases with the coefficient of kurtosis of the log scale error or when the true underlying model is a log normal with constant error variance (on the log scale). For this reason, Manning and Mullahy[120] and Manning et al.[124] suggest, before using GLM, to assess the form of the log-scale residuals of the OLS regression. If the log-scale residuals are heavy-tailed: leptokurtotic (coefficient of kurtosis > 3) or the log-scale error variance (which increases with skewness of the dependent variable) is greater than or equal to one, then log OLS regression (with the appropriate retransformation for heteroskedastic variance) may be preferable to GLM. If, however, the log-scale residuals are both leptokurtotic and heteroskedastic, then the results from both log-OLS regression and GLM should be reported and compared. If the probability density function (pdf) of the raw-scale residuals from one of the GLM estimators with a log link are not bell-shaped or skewed bell-shaped, then OLS-models may be less precise. If the pdf of the raw-scale residuals from GLM are monotonically declining then the appropriate family distribution should be identified using a modified Park test.

Still another problem with GLM is that independent specification of the link and variance functions could lead to bias and estimation inefficiency. Whilst the appropriate variance function may be identified from the modified Park test, one obtains different regression coefficients and inferences on incremental/marginal effects as the link function selected varies. I therefore

considered an extended estimating equations (EEE) version of GLM (also referred to as power-GLM or PGLM) that doesn't require *a priori* specification of the link and variance functions. The PGLM/EEE estimator utilizes a Box-Cox transformation for the link function (to minimize skewness and to gain symmetry in the residual errors). This is defined as:

$$X\beta_k = \begin{cases} \left(\widehat{ADMINCOST}^{\lambda} - 1\right)/\lambda, \text{if } \lambda \neq 0 \\ \ln(\widehat{ADMINCOST}), \text{if } \lambda = 0 \end{cases} \quad (10)$$

Two broad family distributions can be specified: (1) a "power variance" family characterised by: $\text{var}(ADMINCOST; \theta_1, \theta_2) = \theta_1(\widehat{ADMINCOST})^{\theta_2}$ and (2) a "quadratic variance" family characterised by: $\text{var}(ADMINCOST; \theta_1, \theta_2) = \theta_1\left(\widehat{ADMINCOST}\right) + \theta_2(\widehat{ADMINCOST})^2$, where θ_1, θ_2 together index the appropriate variance distribution for the dataset analysed. By *simultaneous* specification of a family of link functions and variance functions, and *joint* estimation of the parameters above, PGLM/EEE is a more flexible and robust estimator especially when no specific link or variance function can be identified. For example, if $\hat{\delta}$ in the GLM estimation is a non-integer then choosing the closest family distribution could lead to efficiency losses[125],[126].

Given the assumptions underlying my analysis and issues with the modelling (estimation) approaches described above, I thought it prudent to validate my results – by comparing the accuracy of the estimates derived from the algorithm to administration-cost estimates reported in other studies.

4.3 Results

4.3.1 Simulations

Figure 4.1 below shows the log-log plot of Pc versus PAc from outputs of the simulations performed for the product sample. This provides further evidence of variation in drug administration costs by the route of administration. The simulations also confirm my expectations that the costs of administering biologics subcutaneously or intramuscularly are mainly from the costs incurred before or after physical administration of a drug – although some deviations (inconsistencies) are evident. Figure 4.1 also shows that for biologics administered subcutaneously or intramuscularly, most of the simulations lie above the 45º line, which equates Pc to PAc. That is to say, a greater proportion of the administration costs for these groups of

biologics are attributable to the proximal costs incurred before or after drug administration. In contrast, the simulations for intravenous biologics fall on either side of the 45º line with the exception of two drugs: trastuzumab and basiliximab. For trastuzumab, most of the simulations fall above the 45º line, which suggests that the associated proximal cost of administering this drug is higher (relative to the physical administration costs). In the case of basiliximab, most the simulations fall below the 45º line indicating that physical administration costs are higher for that drug.

Although Figure 4.1 provides an idea of the general location of the simulated values for Pc and PAc for each product category, the overlap of data points makes it difficult to tell the pattern for each biologic product. The exact percentage of simulations with the ratio PAc/Pc ≥ 1 can, however, be easily computed from the synthesized data. For the intravenous products, the percentage of simulations with the ratio PAc/Pc ≥ 1 is as follows: basiliximab (95%), bevacizumab (35%), cetuximab (28%), infliximab (28%), oftamumab (49%), panitumumab (46%), tocilizumab (49%) and trastuzumab (9%). For the subcutaneous products, this is as follows: adalimumab (2%), canakinumab (7%), certolizumab pegol (1%), denosumab (22%), etanercept (20%), golimumab (22%), omalizaumab (33%) and ustekinumab (1%). For the intramuscular biologics, we have 17% for palivizumab and 25% for interferon beta-1a. In general, we can say that a higher proportion of the administration costs of biologic drugs given subcutaneously or intramuscularly come from the proximal costs incurred before or after drug administration while for intravenous products, costs are incurred in both cost centres.

Figure 4.2 below confirms prior expectations that the costs of administering biologics given subcutaneously or intramuscularly are lower than that of biologics given intravenously. However, the numbers on top of each stacked bar in Figure 4.2 suggests that one has to be cautious when using cost per unit administration to illustrate variations in biologic drug administration costs. Tetteh and Morris[3], for instance, using a selected set of studies offering a comparable scale of measurement for cost per unit administration, report that the administration costs of intravenous biologics appear to be six to eight times that of biologics given subcutaneously or intramuscularly. Differences observed (on the basis of cost per unit administration) disappear, are attenuated or reversed when one considers drug administration costs over a defined period of time. It seems that differences in dosing frequency is, at least, one of the reasons why cost per unit administration may not be an appropriate indicator of the variation in administration costs.

This unfortunately still leaves unanswered the question: by how much do administration costs differ between biologics given intravenously, subcutaneously or intramuscularly?

Figure 4.1: Simulation outputs

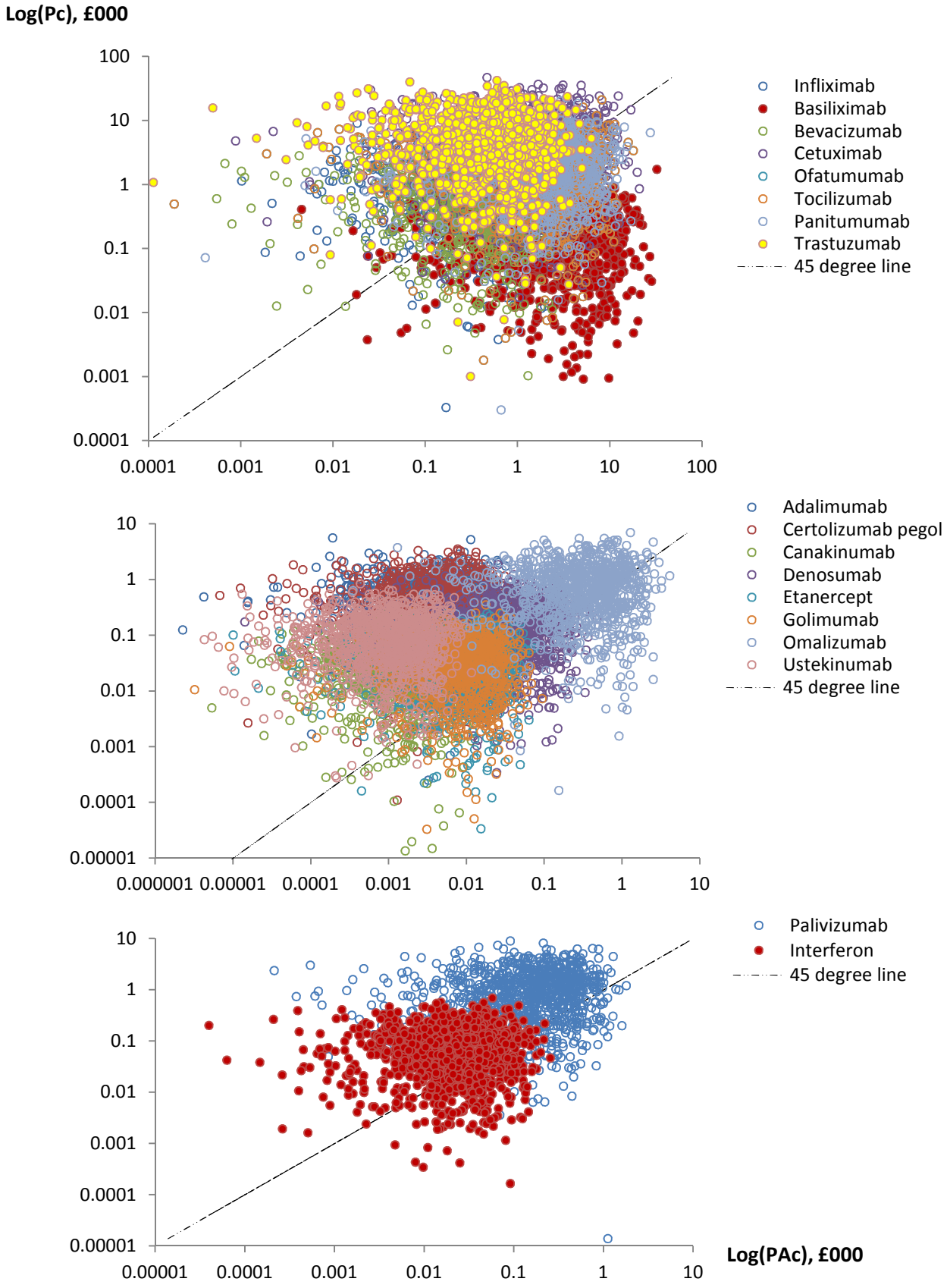
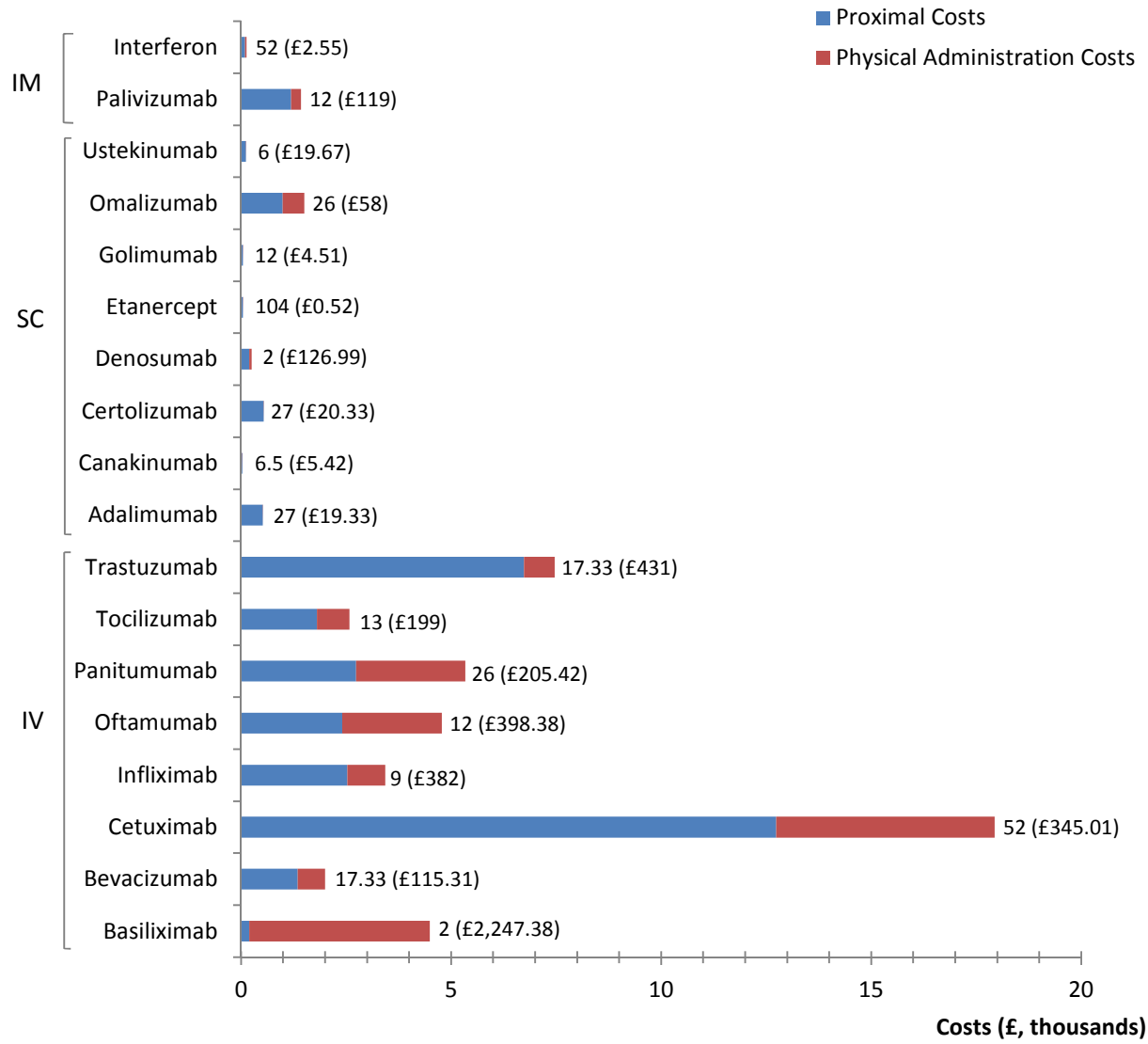


Figure 4.2: Variation in drug administration costs



Notes: IM = intramuscular administration; SC = subcutaneous administration; IV = intravenous administration; values on top of each stacked bar represents DOSFREQ: the number of unit administrations per year; the values in brackets and pound currency refer to the cost per unit administration.

4.3.2 Administration cost algorithm

To answer the question above, I turn to the results shown in Table 4.2. Note that the results are for ln$\widehat{ADMINCOST}$, not $\widehat{ADMINCOST}$. The regression coefficients, nevertheless, carry the desired information as to the link between biologic administration costs and the frequency or route of drug administration. As mentioned earlier, I estimated four nested models A, B, C and D with C as the full unrestricted model. From Table 4.2, model C offers

the best fit with the simulated dataset, having the highest adjusted R^2, lowest sum of squared residuals and the lowest prediction variance, i.e., the square of SE(ln$\widehat{ADMINCOST}$). To ensure robustness, I explored alternative estimators for model C; the results of which are shown in Table 4.3 below.

I found that the raw-scale residuals from a GLM with log-link and gamma family distribution did not exhibit a monotonically declining pdf albeit the residuals were kurtotic (coefficient of kurtosis = 4.3907). The residuals from a standard OLS regression on ADMINCOST were also leptokurtotic (coefficient of kurtosis = 5.080), and remained kurtotic even after of log transformation of ADMINCOST (coefficient = 3.1648). However, whilst tests for normality of OLS residuals indicate non-normal data, one could not reject the null of normal data for the residuals from the log-OLS regression (at the 5% significance level). Variants of the Breusch-Pagan test for heteroskedasticity of the log-OLS residuals showed that, at the 5% significance level, one cannot reject the null of constant variance (lowest *p* value = 0.078). Further, a standard Park test, i.e., an auxiliary regression of the form $\exp(\widehat{\epsilon_i}) = \rho(X)$ indicated that none of the explanatory variables are statistically significant predictors of log-scale residuals. The log-error variance (i.e., the mean squared error from the log-OLS model), however, was less than one – see model C in Table 4.2.

Table 4.2: Identifying the algorithm

Dependent variable: lnADMINCOST				
	Model A	**Model B**	**Model C**	**Model D**
Independent variables	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)
ROUTEADMIN1	-3.2858 (0.57)***	-2.6161 (1.106)**	-3.2073 (0.902)***	—
ROUTEADMIN2	-2.4097 (0.90)**	-2.2784 (1.219)*	-5.1856 (1.335)***	—
DOSFREQ	-0.0046 (0.011)	-0.0025 (0.012)	0.4206 (0.18)**	-0.1022 (0.179)
PRODUCTBUND	—	-0.7395 (1.129)	0.4133 (0.928)	-2.7416 (0.678)***
INDICATN	—	-1.0527 (1.219)	-0.2084 (1.467)	-3.7195 (1.693)**
DOSFREQ2	—	—	-0.00105 (0.0003)***	-0.00075 (0.0004)*
DOSFREQ.INDICATN	—	—	-0.3126 (0.172)*	0.17 (0.183)
Intercept	8.6223 (0.45)***	9.6888 (1.274)***	7.1058 (1.521)***	11.305 (1.685)***
Regression statistics				
R^2 (Adjusted R^2)	0.7188 (0.6586)	0.7376 (0.627)	0.8881 (0.8098)	0.6957 (0.5689)
F-statistic (> c)	11.9313 (> 3.1122)	6.7151 (> 2.9961)	11.3369 (> 3.0717)	5.4864 (> 2.9961)
SSR (N)	17.8023 (18)	16.6714 (18)	7.0858 (18)	19.2688 (18)
SE(ln$\widehat{ADMINCOST}$)	1.1276	1.1787	0.8418	1.2672
F-*ratio* (> c)	3.781 (> 3.478)	6.764 (> 4.1028)	—	10.3161 (> 4.1028)

Notes: $\hat{\beta}_k$ = regression coefficients; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; SE = default standard error; F-statistic = F-test for overall significance of the regression; c = critical value for F-test; SSR = residual sum of squares; N = number of observations; F-*ratio* = F-test for comparison with model C; c = critical value for the F-test; ROUTEADMIN1 = dummy variable for subcutaneous delivery; ROUTEADMIN2 = dummy variable for intramuscular administration; DOSFREQ = dosing frequency; PRODUCTBUND = dummy indicating whether the drug product is sold as a bundled product with some of the equipment and consumables used in drug administration; INDICATN = dummy for type of illness a product is clinically indicated for.

Table 4.3: Alternative estimators for model C

Dependent variable: ADMINCOST or lnADMINCOST				
	GLM (log, Gamma)	**PGLM/EEE (QV)**	**GLM (identity, Gaussian)**	**NLLS**
Independent variables	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)	$\widehat{\beta_k}$(SE)
ROUTEADMIN1	-2.8255 (0.314)***	-2.6618 (0.369)***	-3.2026 (0.363)***	-3.2026 (0.474)***
ROUTEADMIN2	-5.0158 (0.393)***	-4.7495 (0.609)***	-5.2737 (0.509)***	-5.2737 (0.664)***
DOSFREQ	0.4015 (0.042)***	0.3754 (0.059)***	0.428 (0.054)***	0.428 (0.07)***
PRODUCTBUND	0.2364 (0.198)	0.2336 (0.19)	0.404 (0.269)	0.404 (0.35)
INDICATN	-0.2352 (0.145)	-0.2592 (0.133)*	-0.2896 (0.183)	-0.2896 (0.283)
DOSFREQ2	-0.00102 (0.0002)***	-0.00093 (0.0001)***	-0.00106 (0.0002)***	-0.00106 (0.0003)***
DOSFREQ.INDICATN	-0.3004 (0.021)***	-0.2775 (0.053)***	-0.3173 (0.028)***	-0.3173 (0.036)***
Intercept	7.3703 (0.257)***	-0.6008 (0.264)	7.1499 (0.349)***	7.1499 (0.455)***
Regression statistics				
Log pseudo-likelihood (AIC)	-143.3921 (16.377)	—	-16.8641 (2.3182)	—
OD parameter: $\hat{\delta}$(95% CI)	1.589 (1.251, 1.921)	—	—	—
Link parameter: $\hat{\lambda}$(95% CI)	—	0.078 (-0.2361, 0.3921)	—	—
VF parameter: $\widehat{\theta_1}$(95% CI)	—	0.0304 (-0.0038, 0.0646)	—	—
VF parameter: $\widehat{\theta_2}$(95% CI)	—	0.0899 (-0.0038, 0.1836)	—	—
R^2 (Adjusted R^2)	—	—	0.8923 (0.8285)	0.8923 (0.8692)

Notes: $\hat{\beta}_k$ = regression coefficients; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$; SE = heteroskedastic-robust standard error; CI = confidence interval; ROUTEADMIN1 = dummy variable for subcutaneous delivery; GLM = generalized linear modelling; PGLM = Power-GLM; EEE = extended estimating equations; QV = quadratic variance function; NLLS = non-linear least squares; ROUTEADMIN2 = dummy variable for intramuscular administration; DOSFREQ = dosing frequency; PRODUCTBUND = dummy indicating whether the drug product is sold as a bundled product with some of the equipment and consumables used in drug administration; INDICATN = dummy for type of illness a product is clinically indicated for; AIC = Akaike Information Criterion; OD = over-dispersion; VF = variance function.

As noted by Manning and Mullahy[120] and Manning et al.[124], these statistics affect the choice between log-OLS and GLM estimators. A modified Park test following the log-gamma GLM suggests that we cannot reject a Gaussian, poisson, gamma or inverse Gaussian distribution for the GLM variance function. But besides not being able to identify the appropriate GLM link and variance functions, we also know GLM suffers precision losses in the face of heavy-tailed residuals. A joint estimation of the link and variance functions by PGLM/EEE (QV) rejected a gamma GLM with a log link[10]. Note that whilst the regression coefficients from the log-gamma GLM and the PGLM/EEE (QV) are not entirely consistent, the initial coefficient values used in the PGLM/EEE are derived from a gamma GLM with a log-link. Based on the estimated values for $\hat{\lambda}$ and $\widehat{\theta_2}$, the PGLM/EEE (QV) model identifies NLLS or a log-linear model with Gaussian variance function as the best model fit to the synthesized dataset. I implemented the latter using the GLM equivalent of a log-OLS regression as I found that GLM with log-link with a Gaussian variance function failed Pregibon's link test. I believe this is because the natural (canonical) link function for a Gaussian family distribution, especially with small samples, is an identity one – not to mention the precision losses from GLM in the face of heavy-tailed residuals.

Considering the (borderline) statistically-insignificant tests for skewness, kurtosis and heteroskedasticity; the consistent coefficients from the NLLS and GLM equivalent of log-OLS (plus the observation that none of the explanatory variables in model C are statistically-significant predictors of heteroskedastic variance if it exists), I believe that the underlying true model is closer to a log-normal with homoscedastic variance on the log-scale. I focussed my attention therefore on the log-OLS version of model C (column 3 of Table 4.3). But before proceeding, it is important first to remember that the reference category is formulating biologic drugs for intravenous administration. The intercept from model C regression represents lnADMINCOST for an intravenous biologic drug indicated for an acute illness and not sold as a bundled product. This, however, carries no particularly meaningful information as it requires dosing frequency to be zero. There is no point manufacturing and formulating a biologic drug if it is not going to be used. Second, the regression coefficients are semi-

[10] In contrast to Basu and Rathouz[125], I found that the PGLM/EEE with a power variance function failed to converge. Misspecification tests for the PGLM/EEE (QV) suggested a good fit with the data: Pearson correlation between the raw-scale residuals and predicted values was not significantly different from zero, and there was no statistically-significant evidence of systematic patterns in the residuals plotted against predicted values.

elasticities: they represent the *ceteris paribus* percentage change in $\widehat{ADMINCOST}$ given a unit change in the explanatory variables[127],[128],[121]. Wooldridge[121] suggest that for continuous variables, the marginal effects (percentage change in $\widehat{ADMINCOST}$) = 100•$\widehat{\beta_k}$ for a unit change in the explanatory variable X. For dummy variables, the incremental effects (percentage change in $\widehat{ADMINCOST}$) = 100•[exp($\widehat{\beta_k}$) – 1]. In this section, I follow Kennedy's[18] argument that for dummy variables in an OLS regression, the incremental effect is given by: 100•[exp{($\widehat{\beta_k}$) – 0.5(SE($\widehat{\beta_k}$)^2} – 1]. This corrects for small sample bias.

The coefficient for the variable for subcutaneous products has the expected sign; it is statistically significant and suggests that formulating a biologic drug candidate for subcutaneous rather than intravenous delivery, holding all else constant, reduces $\widehat{ADMINCOST}$ by approximately 96.19% (= 100•[exp(-3.2026 – .5(.363^2)) – 1]). The coefficient on the variable for intramuscular products also has the expected sign, is statistically significant and suggests that formulating a biologic drug candidate for intramuscular rather than intravenous delivery will reduce $\widehat{ADMINCOST}$ by approximately 99.55% (= 100•[exp(-5.2737 – .5(.509^2)) – 1]) holding all else constant. These percentages do not mean that intravenous products account for over 99% of the variation in administration costs for my product sample – it only refers to isolated incremental effects of the route-of-administration variables. Note also that the difference in the ceteris paribus effects of the variables for subcutaneous and intramuscular products: $\widehat{\beta_0} - \widehat{\beta_1}$ is statistically significant (p value = 0.0035). That is, there is 'strong' evidence of a reduction in ADMINCOST in switching from subcutaneous to intramuscular formulation at the 5% significance level. Note also that in models A and B, the coefficients for the route-of-administration variables have the expected signs; they are of the same order of magnitude as the coefficients for model C but are not always statistically significant at the 5% level. I considered that route-of-administration variables explain a lot of the variation in lnADMINCOST and ADMINCOST.

The coefficient for dosing frequency has the expected sign and it is statistically significant; but it should not be interpreted in isolation given the quadratic term for dosing frequency and its interaction term (DOSFREQ.INDICATN). The percentage change in $\widehat{ADMINCOST}$ with a unit change in dosing frequency is not 42.8% (= 100•.428). The negative coefficient on the quadratic term for dosing frequency suggests that the marginal

effect of dosing frequency on ln$\widehat{ADMINCOST}$ diminishes with increases in dosing frequency: the relationship between the two is curvilinear. We can write this relationship as follows:

$$\partial(ln\widehat{ADMINCOST})/\partial DOSFREQ = \widehat{\beta_2} - 2\widehat{\beta_5}DOSFREQ - \widehat{\beta_6}INDICATN \qquad (11)$$

For an acute illness, INDICATN = 0 and the percentage change in ADMINCOST from moving from once a year dosing to twice a year is 42.59% (= 100•[.428 – 2(.00106)(1)]). But moving from 100 to 101 unit administrations per year will only increase $\widehat{ADMINCOST}$ by 21.6% (= 100•[.428 – 2(.00106)(100)]). For a chronic illness (INDICATN = 1), the corresponding percentage change in $\widehat{ADMINCOST}$ of moving from one to two unit administrations per year is 10.86% and -10.13% for moving from 100 to 101 unit administrations per year. The latter result reflects the curvilinear relationship between ADMINCOST and dosing frequency. Although the coefficient on dosing frequency suggests that increases in dosing frequency should lead to increases in $\widehat{ADMINCOST}$, the coefficient on its quadratic term means that beyond some positive value of DOSFREQ (a turning point), increases in dosing frequency will be associated with lower drug administration costs.

For biologics indicated for acute illnesses, this turning point is achieved when dosing frequency is approximately 202 (= .428 divided by 2[.00106]) unit administrations per year. This is roughly a four times a week dosing regimen with perhaps a tapering off of dosing frequency due to improved treatment response or concerns about adverse events. For the two biologic drugs indicated for acute illness, their unit administrations per year are far from 202 so we can safely ignore what happens after the turning point. For biologics indicated for chronic illnesses, this turning point is reached when dosing frequency is just over 52 (= [.428 – .3173] divided by 2[.00106]) unit administrations per year, i.e., roughly dosing once weekly. Only one of the chronic biologic drugs within my sample (etanercept) has a dosing frequency that greatly exceeds this turning point; so again one can ignore the turning point knowing the source of this seemingly counterintuitive observation.

Indeed, a test for detecting influential (outlier) observations indicated that only one product (etanercept) had an absolute DFITS value greater than twice the threshold of 1.333. (DFITS is the scaled difference in predicted values of ADMINCOST with and without the j^{th} observation, in this case etanercept.) However, the curvilinear relationship between ADMINCOST and dosing frequency raises an interesting scenario that, given the peculiar

biophysical characteristics and pharmacokinetic profiles of a biologic drug candidate, it is possible to have lower ADMINCOST when dosing frequency is higher relative to some appropriately defined comparator drug.

For the variable indicating a bundled product, I found that whilst it did not reach statistical significance, the regression coefficient has the unexpected sign. The size of the coefficient suggests that, holding all else equal, bundling a biologic product with some of the equipment and consumables used in drug administration increases $\widehat{ADMINCOST}$ by approximately 44.46% (= 100•[exp(.404 – .5(.2693^2)) – 1]). Since one cannot reject the possibility that the coefficient of the variable for bundled products is zero, I sought to explore the dataset to see why I observe this result. It turns out that the there is a non-trivial correlation between the variable for bundled products and the variable for subcutaneous products (correlation coefficient of .6325) and with ROUTEADMIN2 (correlation coefficient of .25). Dropping the route-of-administration variables in model D (Table 2) produced the expected sign for the variable for bundled products and it achieved statistical significance. This suggests that bundling a biologic product with some of the equipment and consumables used for drug administration reduces $\widehat{ADMINCOST}$ by roughly 95% (= 100•[exp(-2.7416 – .5(.678)^2) – 1]) – similar to the ceteris paribus effects of the variables for subcutaneous and intramuscular products. This makes sense as all the products in my sample administered subcutaneously or intramuscularly are, by definition, bundled with some of the equipment and consumables used for administering drugs. Model D, however, suffers from dropping the route-of-administration variables: its adjusted R^2 is the lowest (.5689).

The variable for chronic/acute illness, INDICATN, also shows up with the unexpected sign but it doesn't reach statistical significance in models B and C. This result seems counterintuitive considering the persistence of healthcare resource consumption and expenditures associated with chronic illnesses. In model D (Table 2), however, the variable INDICATN reaches statistical significance. Again, I explored my simulated dataset to find reasons why this might be the case. It turns out that the variable for chronic/acute illness, INDICATN, is negatively correlated with the variable for intramuscular products (correlation coefficient of -.4375) and positively correlated with the variable for subcutaneous products (correlation coefficient of .3162). I speculate that the net outcome of these opposing effects is the reason why INDICATN has the negative sign. In fact, for the subgroup of intramuscular biologics in my sample, there is only one product for acute illness and the other for chronic

illness. There are no subcutaneous biologics indicated for acute illnesses. The apparent counterintuitive result simply reflects the nature of the product sample and simulated dataset. It might not necessarily be observed with a different or updated product sample.

Note that although the variables for bundled products (PRODUCTBUND) and chronic/acute illness (INDICATN) have unexpected signs and individually they may not always reach statistical significance, they are *jointly significant* in the presence of the other explanatory variables in model C. That is to say, a reduced form of the OLS regression model C that excludes these statistically insignificant variables will carry a lot more of *unexplained* variation in ADMINCOST. As indicated by the F-ratios for the comparisons of model A, B and D with model C in Table 4.2, there is no justification for excluding these statistically-insignificant variables.

4.3.3 Validation

As discussed in section 4.2.5, one must proceed with caution when using the algorithm based on the log-OLS model C to predict the administration costs for a biologic drug candidate. Specifically, one needs a smearing factor to minimize prediction errors. I found that Duan's smearing factor for all product categories was 1.204; and the subgroup-specific smearing factors are as follows: 1.0792 for intravenous products, 1.3799 for subcutaneous products and 1.00 for intramuscular products. The smearing factor from Wooldridge's approach was 1.0287. To identify which of the methods of deriving less biased OLS estimates of $\widehat{ADMINCOST}$ is best, I estimated the coefficient of determination (i.e., the square of the correlation coefficient) between the smearing estimates and simulated values for ADMINCOST. This was .822675 for the 'normal theory estimates', .822382 for estimates derived using the subgroup-specific smearing factors and .822675 for the Wooldridge approach. There is therefore not much to choose from the three alternatives but I prefer the subgroup-specific smearing factors as it allows us one to make cost predictions tailored to specific product formulations; and because the residuals from the log-OLS model C exhibit statistically-insignificant heteroskedasticity that is not explained by any of the explanatory variables.

To illustrate how the algorithm can be used to predict administration costs, consider current attempts to reformulate trastuzumab for subcutaneous delivery. This has been made

possible by the feasibility of manufacturing highly-concentrated solutions for monoclonal antibodies and co-formulation with an excipient, recombinant human hyaluronidase that dissolves subcutaneous tissues for rapid drug absorption[129]. Let's assume health-outcome neutrality between subcutaneous and intravenous delivery of trastuzumab, and the same dosing frequency at three-weekly intervals. Recent phase I trials suggest that subcutaneous and intravenous trastuzumab have indeed comparable efficacy and safety profiles [130]. Also subcutaneous trastuzumab will, by definition, be available as a bundled product: it is mostly likely to be sold together with an automated single-use injectable device in place of manual administration with a syringe[129].

Given trastuzumab is indicated for a chronic illness (see Table 4.1), the annual ADMINCOST can be predicted as follows:

$$\widehat{lnADMINCOST_{IV}}$$
$$= 7.1499 - 3.2026(0) - 5.2737(0) + .428(17.33) + .404(0) - .2896(1)$$
$$- .00106(17.33^2) - .3173(17.33)(1)$$

$$\widehat{ADMINCOST_{IV}} = \exp(8.4604) \bullet \hat{\Phi}(= 1.0792) = £5097.99$$

$$\widehat{lnADMINCOST_{SC}}$$
$$= 7.1499 - 3.2026(1) - 5.2737(0) + .428(17.33) + .404(1) - .2896(1)$$
$$- .00106(17.33^2) - .3173(17.33)(1)$$

$$\widehat{ADMINCOST_{SC}} = \exp(5.6618) \bullet \hat{\Phi}(= 1.3799) = £396.94$$

where 1.0792 and 1.3799 are the subgroup-specific smearing factors for intravenous and subcutaneous products respectively. Assuming the same price per given dose, this represents a healthcare delivery cost saving of, at least, £4700 per patient per year.

These results are consistent with Samanta et al.'s[131] report that a 100% switch of 200 patients in England receiving intravenous trastuzumab to its subcutaneous equivalent will generate time and resource cost savings of £271,000 in the hospital setting; £1,200,000 in the community setting and £1,500,000 if patients self-administer subcutaneous trastuzumab at home. The savings arise mainly from reduction in pharmacy technicians' time and nursing

inputs and time spent in the “IV chair”, i.e., being hospitalised to receive an intravenous infusion. The regression-based algorithm provides a conservative estimate of £940,210 if 200 patients fully switch from intravenous to self-administered subcutaneous trastuzumab. Also De Cock et al.[23] report, from a multi-country, multi-centre time and motion study that resources expended in administering intravenous trastuzumab is mainly in the form of reconstitution in the pharmacy and in “infusion initiation” (no changes in number of patients’ visits, blood sampling and physician consultation were expected from the switch to subcutaneous trastuzumab). My results are consistent with this finding; to be specific, the ratio PAc/Pc is greater than or equal to one in only 9% of 1000 simulations for trastuzumab.

4.4 Discussion

The regression-based algorithm developed above allows one to highlight cases where (1) administration costs for a given biologic medicine vary across different clinical indications, by the route of administration and dosing regimens of a given biologic medicine, and (2) how administration costs can be lowered via bundling medicinal products with some of the equipment and consumables used in administering drugs. The essence of the algorithm is better appreciated when one considers the argument by de la Horie[133] that the reason why biologics are “so expensive” is because of complex and costly manufacturing, and the need for frequent administration of high doses to be effective. This suggests that disease management or healthcare delivery costs could be reduced by manufacturing process innovations that reduce the cost-of-goods as well as reformulation steps that reduce the frequency of dosing.

My results suggest that this is possible but then the curvilinear relationship between ADMINCOST and dosing frequency means it is also possible for some biologic drugs (given their peculiar biophysical characteristics and pharmacokinetic profiles) to have lower administration costs even with a higher annual dosing frequency (relative to some other biologic drug). However, for my product sample, an increase in dosing frequency in most cases will be associated with an increase in ADMINCOST albeit at a diminishing marginal rate. But we also know that variables for subcutaneous and intramuscular products are associated with lower administration costs even when dosing frequency remains unchanged.

One could therefore say the key decision factor is with regards to formulating a product for subcutaneous or intramuscular administration relative to intravenous delivery. In line with the argument by Eisenstein[134], a one-size-fits-all approach of formulating biologics for intravenous delivery needs to be reconsidered.

This, however, should not be taken to mean that all biologics should be manufactured for subcutaneous or intramuscular administration. Intravenous administration has other attributes that make it the most appropriate route for administering a drug. This includes the benefit of immediate injection of active drug moieties into the systemic circulation – something that is desired when an immediate treatment or clinical response is needed, for example, with management of epileptic seizures, acute asthmatic attacks or in severe sepsis where peripheral shutdown hinders rapid drug distribution and absorption. Intravenous delivery is also appropriate for products with narrow therapeutic indices as there is less fluctuations of drug levels in blood plasma. Generally speaking, drugs that need 'informed' dosage adjustments (based on accurate measurements of some physiological or biochemical metric) in order to ensure the products deliver positive health benefits (net of safety concerns and risks) are best administered intravenously. Intravenous delivery is also appropriate when a drug cannot be absorbed from the gastrointestinal tract or when a drug cannot be injected into the muscle or other body tissues[135].

Putting aside these clinical reasons, there are manufacturing challenges (issues of technological feasibility) that need to be addressed before the potential efficiency (cost) savings can be realized. Biologics can be difficult to formulate for subcutaneous or intramuscular delivery as this typically involves injection of small volumes of highly concentrated drug solutions through needles with narrow apertures (needles used for subcutaneous often have an aperture of 0.5 inches whilst needles with 1-2 inch gauges are used for intramuscular delivery). The problem is amplified for biologics that need to be given in high doses and/or have limited solubility. Subsequently, attempts to formulate biologic drugs with fragile molecular structures in high concentrations and small volumes could lead to protein aggregation, undesirable viscosity properties and generally 'unstable' drugs that do not retain their biological or biophysical properties. A biologic drug that works well when given in high volumes (because of solubility problems for example) as an intravenous infusion may lose its clinical efficacy or product quality when formulated for subcutaneous or intramuscular delivery in small volumes[136],[134].

What is more if the small volumes of subcutaneous or intramuscular biologics needed to be injected results in an increase in dosing frequency for a biologic product, the net impact will be determined mainly by the opposing effects of the variables for subcutaneous or intramuscular products and the variable for dosing frequency. Note that this effect from frequent subcutaneous or intramuscular injections of small volumes is irrespective of whether the biologic product in question has a longer half-life that, all things being equal, should lead to a lower dosing frequency.

Another reason why subcutaneous or intramuscular formulation might be associated with an increase in dosing frequency or higher doses is that biologics given intramuscularly or subcutaneously will have to go through layers of skin or muscle tissues, and in the process they may be rendered ineffective or fail to reach the desired target sites. That is, if X doses of a drug are needed for the desired clinical outcome, some allowance (W) must be made for the lost or trapped drug doses by administering $X+W$ doses of the drug. In fact, this problem is the reason behind the conduct of clinical trials on reformulated trastuzumab, rituximab and immunoglobulin G with recombinant hyaluronidase to enhance drug absorption following subcutaneous administration[136].

Assuming the net effect of the variables for subcutaneous and intramuscular products and dosing frequency is a reduction in administration costs, this cost saving might be offset by the fact for a given fixed price per dose, a higher dosing frequency increases the acquisition costs for that biologic drug. Again, depending on the trade-off between acquisition and administration costs, the overall impact might therefore not be a reduction in disease management or total healthcare delivery costs. Likewise, for the same dosing frequency, if it costs more to make the reformulated product then to maintain the same price-cost margin, this will lead to a higher product price per dose assuming price demand elasticity remains the same. That said, the regression-based algorithm in these situations should help manufacturers quantify the net impact of the formulation of drug administration, which together with considerations on the impact on acquisition costs will allow them to generate credible estimates of the total healthcare delivery costs for their products and the likelihood that these products will find favourable recommendations from healthcare payers or providers.

It might be argued that the regression-based algorithm is tied to the product sample selected; that different results may be obtained if a different biologic drug sample is used,

perhaps one that has a lot more products that are administered intramuscularly. Besides the observation that biologics are rarely formulated for intramuscular administration, that argument is not specific to the findings here: it is applicable to almost any algorithm that has been developed for one purpose or the other. The regression-based algorithm developed here may not yield the desired predictions in all situations. That aside, there are non-monetary aspects of drug administration that I haven't considered here; for example, needle phobia and patient discomfort; inconvenience, disruption of daily activities (from more frequent drug dosing) and non-compliance issues that might negatively affect patients' health-related quality-of-life. I will argue that it is even possible to have an expanded algorithm that when used to predict the impact on drug formulation choice on healthcare delivery considers both the monetary and non-monetary aspects of drug administration. I suggest that further research is undertaken to evaluate, if possible in monetary terms, the non-monetary attributes of drug administration from both NHS and non-NHS perspectives.

Even then an unanswered question is whether biopharmaceutical manufacturers are faced with adequate incentives to consider alternative drug delivery systems or alter their formulation choices as early as possible. From the perspective of the rational or responsible private biopharmaceutical manufacturer, researching and investing alternative drug delivery systems is worth the time, effort and money if the net present value of that decision is positive (see Chess[137]). That is to say, the discounted present value of the stream of incremental quasi-rents (i.e., additional cash flow revenue minus the cost of goods) that a reformulated biologic product or an alternative drug delivery system is expected to bring should exceed the discounted present value of the incremental costs of developing the alternative drug delivery system or reformulating a product. The discounted present value of expected cash flows will be determined by the favourable recommendations given by healthcare payers and how that allows a product to, at least, maintain its market share or capture incremental demand volumes. The incremental costs involved will include those related to reformulation and manufacturing; toxicology studies and/or clinical trials to establish, at least, no differences in average clinical outcomes.

Within the UK NHS the use of CEA/HTA and current efforts to implement value-based pricing (partly based on estimated cost-effectiveness) could and should provide some incentive for manufacturers to consider the relationship between formulation choices and healthcare delivery costs as early as possible in product development but here I cannot say

anything about whether this, on its own, will get manufacturers to change their tact. I believe this is also worth considering in future research.

4.5 Conclusions

This chapter has evaluated variations in the magnitude of administration costs of biologic drugs, taking care to ensure consistent inclusion of all relevant cost resources. From this, I developed a regression-based algorithm with which manufacturers could possibly predict, during process development, how their choices on manufacturing and formulation may impact on the healthcare delivery costs of their products. My results confirm the general notion that the administration costs of intravenous products is higher than that of products administered subcutaneously or intramuscularly. Based on the algorithm, I found that formulating a biologic drug for subcutaneous or intramuscular delivery relative to intravenous delivery is associated with lower administration costs that holding all else equal should lead to lower total healthcare delivery costs. Increasing the frequency of drug dosing generally will lead to an increase in administration costs but it is possible that this might not always be the case.

There are, however, clinical considerations and manufacturing challenges that might militate against the potential efficiency (cost) savings in administration costs from reformulating biologic products or making use of alternative drug delivery systems. But where and when issues of technological feasibility can be dealt with, (bio)pharmaceutical manufacturers could use the algorithm to quantify the net impact on drug administration costs, which together with considerations on the impact on acquisition costs will allow them to generate credible estimates of the total healthcare delivery costs for their products and the likelihood that these products will find favourable recommendations from healthcare payers.

5 Discrete-choice Modelling of End-user Preferences for Modes of Drug Administration

5.1 Introduction

The systematic review in Chapter 3 (see also Tetteh and Morris[112]) demonstrated the link between manufacturing decisions and the direct monetary costs of drug administration. The key message from this study was: administration of multiple drug doses over time requires different types of medical resources and hence can have a non-trivial impact on the monetary costs of healthcare delivery. This argument, however, does not consider the non-monetary hedonic characteristics (attributes) of administering drugs that are linked to the preferences of patients for different modes of drug administration. That is to say, a full accounting of the societal costs and benefits of resources expended on drug administration should take into account both the direct monetary and indirect non-monetary costs and benefits. This is because a given mode of drug administration that incurs the lowest monetary cost to healthcare payers or providers may incur hidden indirect costs in terms of a mismatch with what is preferred by patients (otherwise healthy people) and healthcare professionals acting on behalf of patients[137].

It is crucial therefore to understand and assess the trade-offs between the non-monetary characteristics (attributes) of drug administration – such that better drugs can be developed and manufactured tailored to patients' preferences. The importance of this is evident from a recent report published by the Knowledge Transfer Network (HealthTech and Medicines) on "[t]he future of high value manufacturing [in the pharmaceutical, biopharmaceutical and medical device sectors] in the UK"[138]. The report identified, among a number of factors, the importance of: (1) early consideration of manufacturing needs, (2) flexible production facilities; (3) reducing cost to the UK NHS and (4) the delivery of better services and improved health outcomes to patients. The last, in particular, focussed on the need for: (1) more stable, effective medicines, (2) novel ways of administering them, and (3) smart [packaging] technologies to monitor usage by patients. A full understanding of the non-monetary aspects of drug administration is necessary if the multiple objectives above are to be achieved.

In this chapter, I aim to evaluate the non-monetary characteristics (attributes) of drug administration, focusing on the preferences of (1) patients, or otherwise healthy people from the UK general public, and (2) healthcare professionals[doctors and nurses] acting on behalf

of patient populations in the US. I do this using discrete choice experiments (DCEs) that belong to a set of methodologies designed for the valuation of goods and services that are not traded on the market.

The key advantages of using a DCE over interviews, focus groups and other in-person surveys are: first, that one can evaluate the strength of preference for different attributes and how they are traded off against one another; and two, one could assign an indirect monetary value to the non-monetary attributes of drug administration. By monetizing preferences for the non-monetary attributes of drug administration, a DCE allows comparison with the direct monetary costs of administering biologic drugs that can be predicted using the algorithm presented in Chapter 4. See also Tetteh and Morris[139]. Incorporating these predicted-cost outputs with preference valuations from the DCE conducted in this chapter into (bio)manufacturing decision-making can be useful in the development and translation of (biologic) drug candidates in R&D pipelines into patient-friendly medicines. Manufacturers will be equipped with the evidence base to make patient-friendly medicines that do not impose a heavy financial burden on healthcare providers and payers.

The application of a DCE here is in accordance with the economic literature on product variety, notably Spence's [140] arguments that the most natural way of evaluating the welfare effects of product differentiation is in "attribute space". That is, if the distribution of end-user preferences for a common set of attributes of a class of goods or services is known for a consuming population, then the (expected) demand for any set of existing or hypothetical goods can be estimated. In contrast to working in "attribute space", conventional welfare analysis in "product space", i.e., evaluating demands for products *as a whole*, do not allow estimation of demands for hypothetical, non-existent or potential goods or services. This application of a DCE also resonates with theoretical and empirical "logit models of product differentiation" and "logit models of [monopolistic] competition" [141],[142],[143],[144],[145].

5.2 Background

5.2.1 Drug administration as a non-marketed discrete good

The starting point is to consider drug administration as a non-marketed good or service. Subject to a random shock to an individual's health stock, drug administration forms only one part of a series of healthcare and health interventions to restore health to pre-illness

levels. Preferences for different modes of drug administration are therefore not an end in itself: they reflect a derived demand for the (incremental) health benefits offered by a given drug. The mode of drug administration simply constitutes a vehicle via which these incremental health benefits are delivered to a patient. Put differently, there is not a free market trading in goods and services *solely* for the purpose of drug administration. Drug administration, on its own, bears the features of a non-marketed good or service. And, DCEs are one of several preference-elicitation methods by which the non-monetary aspects of different modes of drug administration can be evaluated.

A DCE requires, in a survey, elicitation of *stated* preferences in the context of hypothetical scenarios or 'constructed markets' that are meant to mimic what will happen had a real market existed. Just as consumers pick their preferred products from a shelf in a retail store, a DCE presents end-users (consumers) with a sequence of choice scenarios, sets or situations in which they have to choose from among a number of competing mutually-exclusive drugs that are identical apart from the way in which they are administered. In a DCE however, consumers make choices over a collection of "discrete" goods: they choose their preferred good from among a set of lumpy indivisible goods or, a unit or fixed quantity of highly-divisible goods. Each discrete good is defined by a bundle, profile or treatment combination of a common set of attributes that are in different proportions or ratios, i.e., the levels or values for the attributes differ from one good to the other. The attributes and attribute-levels should be those relevant to manufacturing decisions and/or consumer choices.

5.2.2 Product choice in attribute space

Consider the following simplified healthcare market made up of a finite number of pharmaceutical manufacturers on the supply-side and a finite number of end-users (consumers) on the demand-side. This could be patients and/or healthcare professionals, acting on behalf of a given patient population. Each manufacturer supplies drug products that are almost identical except for being differentiated according to their mode of administration to patients. The decision to supply such differentiated drug products is subject to the resources available for production, the state of underlying manufacturing science; each manufacturer's expectation of incremental quasi-rents from doing so (i.e., the additional revenues net of manufacturing costs) and whether the expected volume of quasi-rents exceed any fixed or sunk expenditures on R&D. Following characteristics theory [146; 147], or what Baumol [148] describes as the "abstract product approach", demand-side utility obtained

from consumption of a good or service is derived not from the good or service *per se* but indirectly from the hedonic characteristics (attributes) embodied by that good or service. In this market, therefore, demand for a drug product (which is in effect demand for modes of drug administration) can be considered as a derived demand for a bundle of attributes of drug administration.

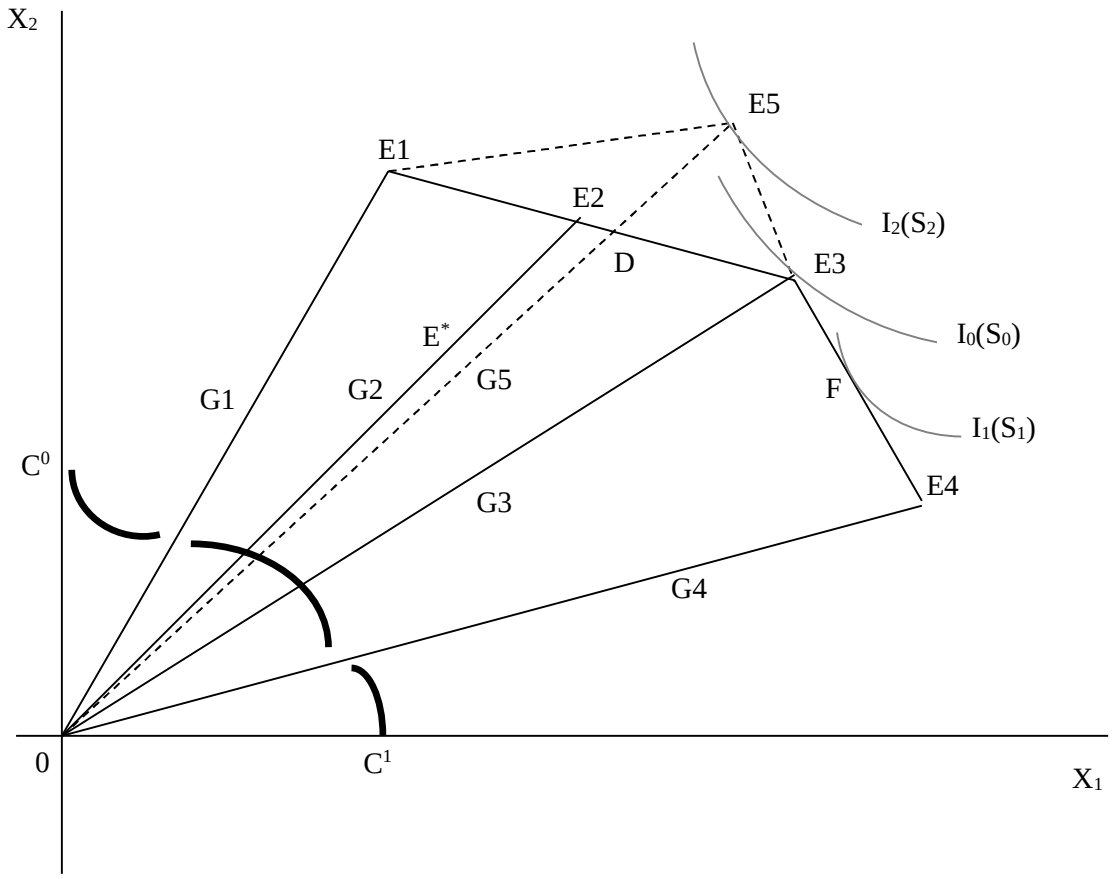


Figure 5.1: Manufacturing challenges and product diversity

Source: Authors; Lancaster [146; 147].
Notes: The upward sloping nature of the rays from the origin reflects the complementary combination of attributes and attribute-levels in defining a given product. C^0C^1 is a discontinuous product differentiation curve determined by the resources available to manufacturers, the state of underlying manufacturing science and the expected incremental quasi-rents from product differentiation.

Initially, in Figure 5.1 above, four manufacturers can possibly supply four classes of highly-divisible drug products G1, G2, G3 and G4, each defined by two attributes X_1 and X_2 in a fixed proportion or ratio (the attribute-levels will be points on the *y*- and *x*-axis). Each class contains drug products defined by various possible combinations of attributes X_1 and X_2, and their levels. These "treatment combinations" of attributes and attribute-levels (or profiles of the products) are, for simplicity, represented by the four rays from the origin. Each ray can

be thought of as the output of a transformation matrix that turns attributes into products, and vice versa. Depending on the resources available to the end-user, i.e., the residual income or financing available after expenditure on a composite of all other healthcare goods and services, he or she will seek to maximize what is often assumed to be a quasi-concave indirect utility function with respect to a vector of the attributes and attribute-levels of drug administration. Given the prevailing prices of drug products in classes G1, G2, G3 and G4, the line $E_1E_2E_3E_4$ defines the "attributes efficiency frontier": the maximum possible combination of attributes and attribute-levels (collection of drug products) that can be afforded within healthcare professionals' funding budget or patients' income constraints. I assume here that healthcare professionals acts as fund-holders (payers): they have financial responsibility over the allocation and use of resources for healthcare provision.

The choice of any end-user then will be determined by the point where the highest indifference curve achievable is tangential to the attributes frontier; for example, point E_3 for an end-user whose indifference curve is I_0. The slope of the indifference curve at this point is S_0 and it represents the marginal rate of substitution between the attributes $X1$ and $X2$. This captures the end-user's willingness to trade off an additional unit of $X1$ for $X2$ in order to maintain the same level of utility. This can be interpreted as the marginal-willingness-to-pay (MWTP) for $X1$ relative to $X2$. An end-user whose highest possible indifference curve (I_1) is tangential to the attributes frontier at point F will prefer a drug product that is some combination of the profiles in classes G3 and G4. However, there is no manufacturer supplying products that contains the profile at point F. As a satisficing option, the end-user may choose to consume at some times products with profile E3 and at other times products with profile E4. The end-user, be it a patient or healthcare professional, switches between products to minimize the gross welfare loss of not having what they want. Note that the assumption of highly-divisible drug products is necessary for X_1 and X_2 to be thought of as representing attributes per unit price of a good. In the more realistic case of indivisible goods or services, price will have to be considered *separately* as an attribute.

In either case, substitution between attributes and attribute-levels will depend on the end-user's MWTP for attributes and how large price changes are. For example, if the price of the product at E2 is to increase or preferences are to change such that the feasible combination of attributes that the end-user can afford is E^*, the new attribute frontier will be $E_1E_3E_4$. That is, some combination of the profiles of products in classes G1 and G3 now dominates those in class G2. On the other hand, and assuming easily surmountable market entry barriers, a new manufacturer introducing a new product belong to a new class G5 at

point D will not, in all practical terms, alter the attribute frontier: combinations of the attributes and attribute-levels in classes G1, G2, G3 and G4 will not dominate combinations including class G5. If, however, the price of products in G5 is low enough, such that the feasible combination of attributes is pushed outward to E_5, then we have a new frontier ($E_1E_5E_3E_4$) where products belonging to class G2 drop out of the preferred bundle corresponding to the indifference curve I_1. The introduction of G5 in this case leads to cannibalization of quasi-rents enjoyed by the manufacturer of products in G2. Depending on the distribution of end-user preferences, introduction of G5 may or may not capture quasi-rents (market shares) enjoyed by products in G1, G3 and G4.

The above, however, assumes that production possibilities (in part determined by the state of the underlying manufacturing science) allow manufacturers to supply drug products in classes G1 to G5 at all levels of attributes *X*1 and *X*2. Now imagine the case of limited production possibilities depicted by the discontinuous product differentiation curve (PDC) labelled C^0C^1. The feasible class of products that can be supplied on the market will be those that belong to classes G2, G3 and G5 but supplied at attribute-levels that fall below or on the PDC. This state of affairs will be characterised by larger pockets of welfare loss to end-users. What is clear here is: given limited resources available to manufacturers, and the need to minimize gross consumer welfare losses, it is crucial that manufacturers have some knowledge of the distribution of end-users' preferences in order for them to supply the classes of drug products (differentiated by their mode of administration) that matches closely what the average representative end-user wants or prefers.

5.2.3 Attribute space to logit demand systems

The framework above accommodates: (1) cases where end-user choices are rational, deterministic and derived from unrestricted (compensatory) substitution between attributes and attribute-levels; and (2) cases where end-user preferences could be any weighted sum of their beliefs about the (incremental) net benefits of a good, formed out of emotions or application of rational or irrational, compensatory or non-compensatory/ lexicographic decision rules[11]. This stochastic nature of choice preferences can be conveniently captured by theorizing that the indirect utility (U_{sj}^*) an end-user, *s*, derives from choosing alternative *j* from among a set of *J* products has: (1) a systematic, explainable or observable component,

[11] Non-compensatory [lexicographic] choices are borne out of hierarchical preference ordering involving no substitution between "dominant" and "dominated" attributes and attribute-levels.

V_{sj} that is a function of the set of attributes and (2) a random unexplainable error term, ϵ_{sj} [149]. We can thus write the following utility function that is linear in parameters and linear in attribute levels:

$$\begin{aligned} \mathrm{U}_{\mathrm{sj}}^{*} &= \mathrm{V}_{\mathrm{sj}}\left(\boldsymbol{\beta}_{\mathrm{jk}}\mathbf{X}_{\mathrm{jk}}\right) + \epsilon_{\mathrm{sj}} \\ \boldsymbol{\beta}_{\mathrm{jk}}\mathbf{X}_{\mathrm{jk}} &= \beta_{\mathrm{k}}'\mathrm{X}_{\mathrm{k}}' + \beta_{\mathrm{p}}\mathrm{C} + \sum_{\mathrm{j}=1}^{\mathrm{J}-1}\mathrm{ASC}_{\mathrm{j}} \end{aligned} \tag{12}$$

where $\boldsymbol{X}_{jk}$ is a vector of attribute-levels decomposed into X_{jk}', a vector of non-monetary attribute-levels and C_{jp}, the price/cost associated with each alternative product. $\boldsymbol{\beta}_{jk}$is a vector of preference coefficients, decomposed into β_{jk}', a vector of coefficients for the non-monetary attributes and β_{jp}, coefficient for the price/cost attribute. The random error term (ϵ_{sj}) refers to the influence of unobserved or unmeasured factors whilst the alternative-specific-constant ASC_j captures any peculiar effects of each alternative product that is not reflected in the attributes. ($\sum_{\mathrm{j}=1}^{\mathrm{J}-1}\mathrm{ASC}_{\mathrm{j}}$ may be considered as the mean of ϵ_{sj}.)

The focus here, however, is consumption of a "discrete good" in fixed quantities. We want to know whether drug products differentiated by their mode of administration will be chosen or not. There is no question of how many or how much.

Given the interest in choices made on the extensive margin, the loci of preferred drug products in attribute space (Figure 1) can be equivalently described by the distribution of choice probabilities for different modes of drug administration. Conditional on knowing the vector $\boldsymbol{\beta}_{jk}$, the probability (P) that product $j1$ will be chosen above the other $J - 1$ discrete products, in each choice situation (n), can be estimated using the "mother" multinomial logit (MNL) model [150] as follows:

$$\mathrm{P}_{1\mathrm{ns}}(\mathrm{y}_1 = 1) = \mathrm{P}_{1\mathrm{ns}}\left(U_1^{*} > U_j^{*}\right) = \frac{\exp(\mu[\boldsymbol{\beta}_{1\mathrm{k}}\mathbf{X}_{1\mathrm{k}}])}{\sum_{\mathrm{j}=1}^{\mathrm{J}}\exp(\mu[\boldsymbol{\beta}_{\mathrm{jk}}\mathbf{X}_{\mathrm{jk}}])} \tag{13}$$

$$f\left(\epsilon_{\mathrm{sj}}\right) = \exp\left(-\epsilon_{\mathrm{sj}}\right)\exp(-\exp(\mu\epsilon_{\mathrm{sj}}));\ \mathrm{F}(\epsilon_{\mathrm{s}}) = \prod_{\mathrm{j}=1}^{\mathrm{J}}\exp(-\exp(\mu\epsilon_{\mathrm{sj}}))$$

where y denotes the choice data collected such that $\mathrm{y}_1 = 1$ if product $j1$ is selected from among J products or options; $f(\epsilon_{sj})$ refers to a Gumbel probability density function for an independent and identically distributed (IID) error term and $F(\epsilon_{sj})$ is the corresponding

cumulative density function. μ is a positive scale parameter that is inversely related to the error variance (σ_ϵ^2) and, for Gumbel-distributed errors, $\sigma_\epsilon^2 = \pi^2/6\mu^2$.

Note that the error terms are specific to each choice dataset, and for that reason μ in the "mother" MNL model is usually normalized to one. The normalized IID error terms together with fixed preference coefficients yields the so-called independence from [ir]relevant alternatives (IIA) assumption – which suggests the ratio of choice probabilities are independent of the inclusion or omission of other products.

The slope of the indifference curve at the locus of product choice in attribute space, for a given end-user, can be described by their MWTP. This is estimated as the ratio of attribute-coefficients and the price/cost coefficient:

$$\mathrm{MWTP}_s = -\frac{\partial(\mathrm{V_{sj}}|\mathrm{X_{jk}})/\partial \mathrm{X_{jk}}}{\partial(\mathrm{V_{sj}}|\mathrm{C_{jp}})/\partial \mathrm{C_{jp}}} = -\beta_{\mathrm{jk}}/\beta_{\mathrm{jp}} \tag{14}$$

Following the setup in Figure 1, one could, in a survey or some referendum, elicit and collect repeated preferences for a collection of products (profiles of attributes and attribute-levels) over a sequence of choice situations – so as to maximize the pool of information that can be collected from a single end-user. Given a sample population of end-users (*S*), and a number of choice situations or sets (*N*) faced by each end-user, the parameters of equations (13) and (14) can be identified by maximizing the sample log-likelihood function ($\mathrm{LL_S}$) with respect to the estimable coefficients. In the case of the "mother" MNL, the likelihood that the choices collected in a DCE survey are explained by the model is given by:

$$\underset{\widetilde{\boldsymbol{\beta}}}{\text{maximizing}}\ \mathrm{LL_S} = \sum_{s=1}^{S}\sum_{n=1}^{N}\sum_{j=1}^{J} \mathrm{y_{jns}} \log \mathrm{P_{jns}}(\widetilde{\boldsymbol{\beta}}, \mathbf{X}) \tag{15}$$

where $\widetilde{\boldsymbol{\beta}}$ refers to prior (expected) coefficient values and all other terms are as previously defined. In what follows, I work with two separate sample populations of end-users: people from the UK general public and healthcare professionals in the US.

5.3. Data and methods

5.3.1 Specification of the attribute space

5.3.1.1 Sample UK general public

In conducting the DCE, I first set out to identify a common set of relevant attributes and attribute-levels for different modes of drug administration. Table 5.1 on page 114 shows the selected set of attributes, definitions of these attributes and their levels for the sample of people from the UK general public. This generic set of attributes and attribute-levels were based on a selective review of literature on preferences for modes of drug administration[151],[152],[153],[154],[155]. They were also informed by two systematic reviews that reported that, excluding wrong-timing errors, there is a higher rate of medication administration errors for intravenous drugs (53.3%) compared to all other routes of drug administration (20.1%). On its own, intravenous drug therapy is associated with a 0.73 overall (Bayesian) probability of making at least one error (95% credible interval [CrI]: 0.54 – 0.90). This was thought to be largely due to errors made at the drug reconstitution stage; hence the use of pre-prepared injections could reduce administration error probability to 0.17 (95% CrI: 0.09 – 0.27)[156],[157].

Ideally, one would want to supplement this literature review with interviews and/or focus group discussions involving end-users. Given the time and resources available for this research, it was not possible to apply these qualitative methods – which are of most value where there is a lack or dearth of existing (grey) literature. I therefore make no claim here that the selected set of attributes and attribute-levels are "exhaustive" of all characteristics of all possible modes of drug administration. I believe, however, that the selected attributes and attribute-levels in Table 5.1 are relevant and suited for investigating the gross welfare benefits (consumer surplus) from manufacturing patient-friendly medicines.

5.3.1.2 Sample US healthcare professionals

Table 5.2 on page 115 shows the generic set of attributes used for evaluating the preferences of US healthcare professionals for different modes of drug administration. The attribute "risk of non-compliance" can be thought of as a composite measure of how a given mode of drug delivery disrupts patients' daily activities; the incidence of adverse events specific to the mode of drug delivery (and separate from that of the drug molecule itself) as well as any other factors that might negatively affect treatment compliance, for instance,

disutility of pain at the site of drug administration, time and travel costs of accessing healthcare or the absence of insurance cover for medical expenses.

In the econometric analyses (see section 5.3.5), I changed the attribute "dosing frequency" into a continuous variable that describes the number of unit administrations of a drug. This was to allow more flexibility in estimating the gross consumer welfare benefits derived from the decisions and actions taken by healthcare professionals on behalf of patients. For the same reason, I translated the cost attribute into a continuous variable, setting an upper limit of $20,000. This is not an arbitrary figure: the levels selected for the cost attribute was meant to mimic the distribution of drug administration costs in Chapter 4, Figure 4.2 on page 89. See also Tetteh and Morris [139]. The maximum cost limit is consistent with Farroni et al. [158] who report that clinical charges for using a room and administering azacitidine in the office (i.e., a clinical setting) ranges from $300-$500 per injection and these costs/charges per patient per year for a full treatment course of six cycles ranges from $12,600-$21,000. I chose $20,000 as the upper limit since charges do not always match costs.

5.3.2 Survey development

Considering the mode of administration is simply a vehicle for delivering the health benefits offered by a drug to patients, I opted to specify the alternative goods as drugs that are identical in every aspect apart from the manner in which they are administered to patients. I used a "forced-choice" format of presenting survey respondents with two unlabelled drugs *A* and *B*. I did not include a "none" or "opt-out" alternative as I found it difficult to imagine that people or healthcare professionals will choose not to have a clinically-beneficial drug simply because the way in which the drug is administered is not what they prefer.

Having identified the most relevant attributes and attribute-levels, I proceeded to develop "efficiency choice" experimental designs for the development of the DCE survey questionnaires. The experimental design is basically an arrangement of *J* alternative products and a sequence of *N* choice sets, where each alternative product is defined by a profile or treatment combination of attributes and attribute-levels. The experimental design embodies the transformation matrix for turning products into bundles of attributes and bundles of attributes into products – such that the speculative choice preferences depicted in Figure 5.1 can be estimated empirically.

Table 5.1: Attributes, definitions and attribute-levels

Attributes	Definitions	Levels
Method of drug administration	This attribute refers to the route by which therapeutically-active drug products are physically administered into a patient. The attribute-levels include all other "needle-free" methods of drug administration to capture the preferences of patients who desire oral drug delivery and/or have a fear of needles.	1. Intravenous delivery 2. Subcutaneous delivery 3. Intramuscular delivery 4. Needle-free delivery
Dosing frequency	This attribute refers to the frequency of administering a drug for a single full course of treatment. Dosing frequency associated with repeated treatments should not be considered.	1. Once every six months 2. Once every month 3. Once every week 4. Once every day
Setting	This attribute refers to place (clinical and non-clinical settings) where a given drug is administered. Clinical settings include, for example, hospitals, outpatient clinics, care homes, offices of general practitioners/physicians etc. Non-clinical settings include home, schools and other public places.	1. Clinical 2. Non-clinical + self-administration* 3. Non-clinical + supervision*
Disruption to daily activities	This attribute refers to how a given method of drug administration or dosing frequency disrupts the daily activities of patients. Disruptions could be due to, for example, repeated venepuncture and, in the extreme, immobility (hospitalization for the sole purpose of drug administration).	1. None 2. Moderate but manageable 3. Moderate but I can't cope 4. Severe
Risk of adverse events	This attribute refers to features of drug administration that might cause discomfort or injury to patients or health-staff administering drugs. This could be local or generalized adverse events such as indurations; damage to nerves and blood vessels; abscess formation around the sites of injection etc. This is separate from side-effects of the drug molecule itself.	1. None 2. Moderate 3. Severe
Price/Cost	This attribute refers to the additional time and travel costs borne out-of-pocket by the patient each time they have to take or their medicines or it has to be given to them by health workers.	1. £0 2. £10 3. £50 4. £100

Notes: * This refers to the situation where people, if properly trained, could self-administer the drug in a non-clinical setting; or otherwise, their medications will have to be delivered to them under the supervision of qualified healthcare professional, for example, a community or district nurse. Given this set of attributes and attribute-levels, we have a full factorial of 2,304 (= $4^4 3^2$) possible profiles or treatment combinations.

Table 5.2: Attributes, definitions and attribute-levels

Attributes	Definitions	Levels
Method of drug administration	This attribute refers to the route by which therapeutically-active drug products are physically administered into a patient. The attribute-levels include "all other needle-free" methods of drug administration to capture the preferences of patients who desire oral drug delivery and/or have a fear of needles.	1. Intravenous delivery 2. Subcutaneous delivery 3. Intramuscular delivery 4. Needle-free delivery
Dosing frequency	This attribute refers to the frequency of administering a drug for a single full course of treatment. Dosing frequency associated with repeated treatments should not be considered.	1. Once every six months 2. Once every month 3. Once every week 4. Once every day
Setting	This attribute refers to place (clinical and non-clinical settings) where a given drug is administered. Clinical settings include, for example, hospitals, outpatient clinics, care homes, offices of general practitioners/physicians etc. Non-clinical settings include home, schools and other public places.	1. Clinical 2. Non-clinical + self-administration* 3. Non-clinical + supervision*
Risk of non-compliance	This attribute refers to any potential threats to medication compliance or adherence due to a given mode or method of drug administration and/or recommended dosing regimen. This is separate from non-compliance due to the safety profile of the drug molecule.	1. None 2. Moderate 3. Severe
Risk of medication errors	This attribute refers to the incidence of common errors of drug administration such as drug preparation and dosing errors; substitution errors (i.e., giving the wrong drug to the wrong patient); violation of sterile conditions when drawing up a drug; cuts in glass ampoules and injection of minute shards of glass with the drug etc.	1. None 2. Moderate 3. Severe
Price/Cost	This attribute refers to the additional resource costs (per patient per full treatment course) incurred in administering drugs to patients.	1. $200 2. $1000 3. $3000 4. Over $3000

Notes: * This refers to the situation where people, if properly trained, could self-administer the drug in a non-clinical setting; or otherwise, their medications will have to be delivered to them under the supervision of qualified health worker, for example, a community or district nurse. For this set of attributes and attribute-levels, we have a full factorial of 1,728 ($= 4^3 3^3$) possible profiles or treatment combinations.

5.3.2.1 Choosing an experimental design

For any DCE, one wants to choose experimental designs that offer some guarantee of finding statistically reliable estimates of $\boldsymbol{\beta}$ at which the first-order Jacobian derivatives of the log-likelihood function of the econometric model specified (to analyse the choice data yet to be collected) is equal to zero. With appropriate coding, these experimental designs can be described by what is called a Fisher information matrix (IM), which is equivalent to the expected negative Hessian matrix of the second-order derivatives of the log-likelihood function of the model chosen for estimation. The inverse of the Fisher IM gives the asymptotic variance-covariance (AVC) matrix denoted by $\boldsymbol{\Omega}$. The off-diagonal elements of the AVC matrix give the parameter covariances, cov($\widetilde{\boldsymbol{\beta}}$). The diagonals contain the expected parameter variance, var($\widetilde{\boldsymbol{\beta}}$) – the square roots of which give the expected standard errors, se($\widetilde{\boldsymbol{\beta}}$).

If all survey respondents observe the same sequence of choice sets, the Fisher IM and AVC matrix can be compactly defined as follows:

$$\mathrm{IM_S}(\boldsymbol{\beta},\mathbf{X}) = -\mathrm{E}\left(\mathrm{S}\cdot\frac{\partial^2 \mathrm{LL}_1(\boldsymbol{\beta},\mathbf{X},\mathbf{y})}{\partial\beta\,\partial\beta'}\right) = \mathbf{X}'\mathbf{P}\mathbf{X} \tag{16}$$

$$\boldsymbol{\Omega} = -[\mathrm{IM_S}(\boldsymbol{\beta},\mathbf{X})]^{-1} = -\frac{1}{\mathrm{S}}\left[\frac{\partial^2 \mathrm{LL}_1(\boldsymbol{\beta},\mathbf{X},\mathbf{y})}{\partial\beta\,\partial\beta'}\right]^{-1}$$

where **X** is the experimental-design matrix of attributes and attribute-levels[12]; **X**' is the transpose of **X**; **P** is a matrix containing the expected choice probabilities; and LL_1 is the log-likelihood for a single average respondent. A design is said to be statistically more efficient than a competing design if: (1) its AVC matrix yields lower expected standard errors, se($\widetilde{\boldsymbol{\beta}}$); and/or (2) with the same choice data, it yields lower estimates of standard errors, se($\widehat{\boldsymbol{\beta}}$). The former is often assessed by a D-efficiency criterion, which can be expressed in absolute or relative terms. Relative D-efficiency of a design is computed as the ratio of the determinant of its AVC matrix design to that of the AVC matrix of some optimal hypothetical design with the largest possible determinant. Identifying a truly efficient experimental design therefore requires prior knowledge of the unknown $\boldsymbol{\beta}$ parameters. So for analytical convenience, most

[12] The experimental design **X** is constructed to have any number of rows ("runs") but this should be greater than or equal to the number of parameters; and the number of columns in **X** should be equal to the total number of attribute-levels. The maximum number for runs is given by the product of the number of alternatives and the number of choice sets.

researchers assume a vector of expected coefficients $\tilde{\boldsymbol{\beta}}$ comprising of zeroes[13] [159],[160],[161].

The experimental designs I used for my DCEs were created from fractional-factorial designs for estimating only *main effects* of the attribute-levels. This was done in SAS v. 9.3, using a set of macros and programming codes written by Kuhfeld[162],[163] as follows. I first used the macro "MktRuns" to gain some insights as to the appropriate number of runs, i.e., the sizes of candidate set-designs I could use. The "MktRuns" macro suggested (among others) the following sizes: 48 runs (= 24 choice sets), 72 runs (= 36 choice sets) and 144 runs (= 72 choice sets). I then used the macro "MkTex" to create corresponding candidate set-designs in 48, 72 and 144 runs. Using the macro "ChoicEff", I identified and evaluated the statistical efficiency of the "best" experimental design containing 24 choice sets and drawn from these candidate set-designs. To test the integrity of the experimental designs above (since I had no prior information on the $\boldsymbol{\beta}$ vector), I merged them with simulated discrete-choices using the macro "MktMerge". Given the artificial dataset created, I estimated a "mother" MNL model using the macro "phChoice" and SAS PHREG procedure.

However, there is a trade-off here: an experimental design with the highest possible D-efficiency may impose greater cognitive burden (task complexity) on survey respondents. One has to balance a desire for near-optimal designs with the possibility of collecting irrational or inconsistent choices. I therefore, using the macro "MktBlock", partitioned the chosen experimental design into two versions – such that each block version contained a sequence of 12 randomly allocated choice tasks.

5.3.2.2 Sample UK general public

I found a 24 choice-set experimental design, drawn from a fractional-factorial candidate set-design with 48 runs. This design had a relative D-efficiency of 67.54% prior to excluding choice sets with "dominated alternatives". Excluding these choice sets yielded a relative D-efficiency of approximately 65%. The relative D-efficiency of another 24 choice-set design, this time drawn from a candidate set-design with 72 runs was 57.35%. This design

[13] The assumption effectively transforms the experimental design from one intended to estimate a non-linear model into one intended for a linear econometric model. These are also referred to as utility-neutral designs as they test a null hypothesis of no difference in the effects of attribute-levels on choice probabilities. A possible improvement will (perhaps in a Bayesian framework) specify a vector of fixed (random) priors close to $\tilde{\boldsymbol{\beta}}$. The benefit from this is a reduced need for larger sample sizes of respondents. On the other hand, specifying the wrong fixed priors (or wrong distributions for the random priors) could lead to significant losses in statistical efficiency. Hence, most researchers take the "less risky strategy" of assuming $\tilde{\boldsymbol{\beta}} = 0$.

had no choice sets with "dominated alternatives". I also found a 24 choice-set experimental design drawn from candidate set-design with 144 runs. This design had a relative D-efficiency of 52.67% prior to excluding choice sets with "dominated alternatives". Excluding these dominated choice sets yielded a relative D-efficiency of 52.42%.

Compared with the other competing 24 choice-set designs developed from candidate set-designs with 72 and 144 runs, the design developed from the candidate with 48 runs, produced the lowest estimates of standard errors over *all* attribute-coefficients for the same simulated choice data. It also had the highest number of statistically-significant attribute-coefficients[14]. I therefore chose this design for the UK survey questionnaire.

5.3.2.3 Sample US healthcare professionals

For the questionnaire(s) aimed at US healthcare professionals, I found a 24 choice-set experimental design drawn from a candidate set-design with 48 runs. This design had a relative D-efficiency of 69.04% and did not appear to contain choice sets with "dominated alternatives". Another 24 choice-set experimental design, this time drawn from a candidate set-design with 72 runs had a relative D-efficiency of 59.59% and contained no choice sets with "dominated alternatives". A similar 24 choice-set design drawn from a candidate set-design with 144 runs had a relative D-efficiency of 52.37%. Excluding choice sets with "dominated alternatives" yielded a relative D-efficiency of 52.24%.

I found that the 24 choice-set design drawn from a candidate set-design with 144 runs, and having the lowest relative D-efficiency (52.24%), produced the lowest se($\widehat{\boldsymbol{\beta}}$) over *all* the parameters with the same simulated choice data. However, the number of statistically-significant parameter estimates ($\widehat{\boldsymbol{\beta}}$) was the same as that of the design drawn from the candidate set-design with 48 runs. On balance, I chose the design drawn from the candidate set-design with 144 runs for the US survey questionnaire.

5.3.3 Survey administration

From the blocked experimental designs above, I developed two draft versions of the survey questionnaire. Each questionnaire was split into three sections. The first section

[14] Note that the simulated discrete-choices were not derived from known utility functions or a known $\boldsymbol{\beta}$ vector. My interests therefore are in the robustness of the attribute-coefficients ($\widehat{\boldsymbol{\beta}}$) and not how $\widehat{\boldsymbol{\beta}}$ closely approximates prior or "true" values of $\boldsymbol{\beta}$. The lower standard errors and high number of statistically-significant coefficients, obtained from the same MNL model estimated with the same simulated dataset, confirms the efficiency gain from the chosen experimental design.

provided a preamble with information about the purpose of the study and the hypothetical "constructed" context in which respondents had to make their choices. It also provided descriptions of the attributes and attribute-levels as well as an example of a completed choice set as a guide for the survey respondents (see Figure 5.2 on page 121). The second section contained the actual sequence of 12 choice questions or sets; and the third section collected anonymized information on individual respondents' characteristics. The anonymized format of the questionnaires meant no ethical approval was required prior to administering the surveys.

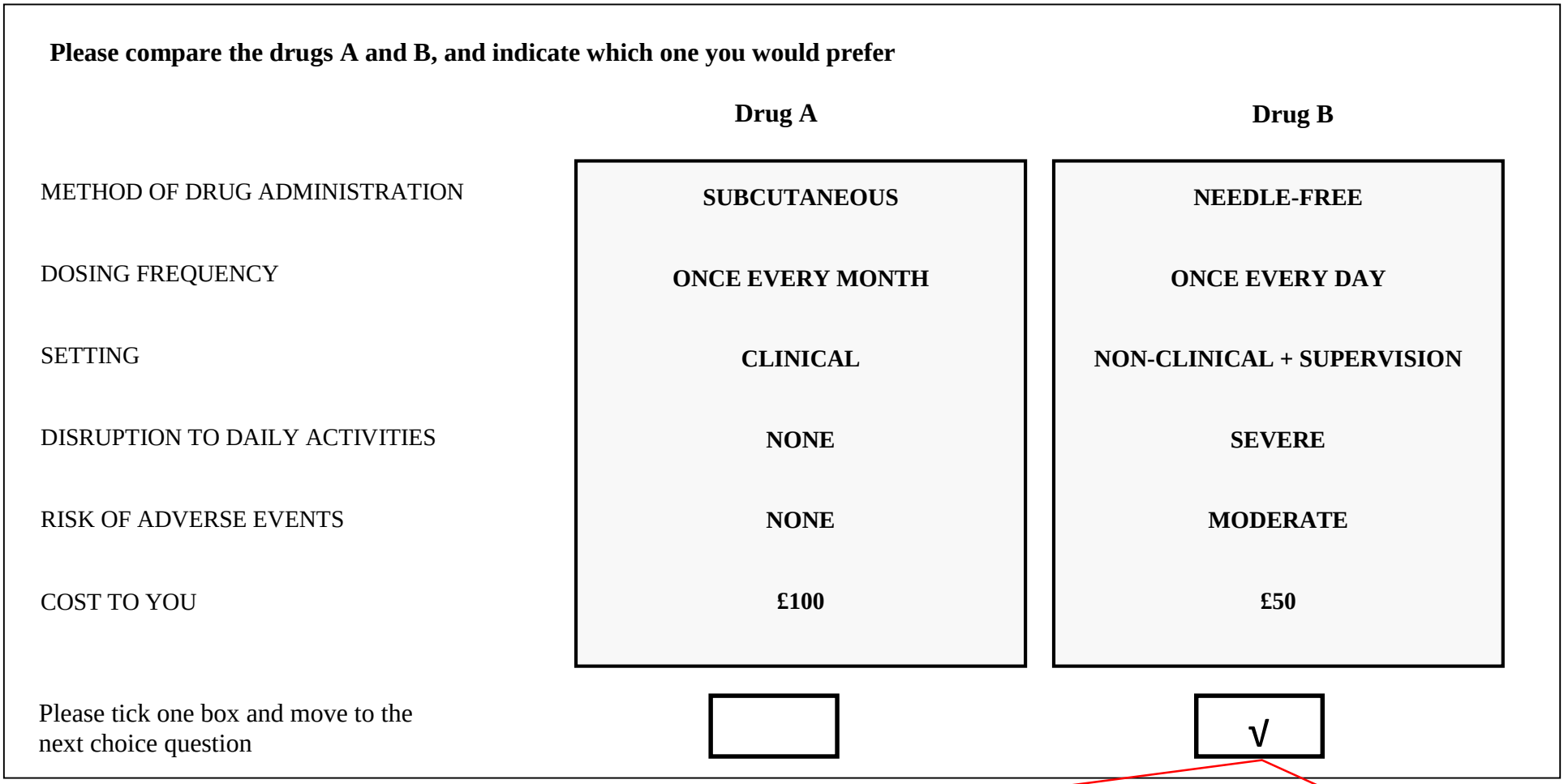
Before sending the questionnaires out, I carried out a small-scale informal pilot of the draft versions of the questionnaires with no more than 5 people from the UK general public (given the time and resources available for this study). The recipients of these questionnaires were asked to check the wording of the questionnaire; to ensure that the instructions were clear and to identify what might be perceived as implausible combination of attributes and attribute-levels. The respondents found all combinations of attributes and attribute-levels plausible, although some combinations may not be technologically feasible (given the current state of manufacturing or formulation science). It took, on average, 15-20 minutes to complete each block version of the questionnaire. Following the pilot phase, I made small wording changes to the questionnaire to improve clarity. With the help of a commercial vendor (Survey Monkey), the questionnaires were administered online to the intended sample populations (i.e., people in the UK general public and healthcare professionals [doctors and nurses] in the US selected from Survey Monkey's database). It took roughly two weeks for the vendor to complete the web-based surveys.

5.3.4 Sample size

According to McFadden[164], "…sample sizes which yield less than thirty responses per alternative [good] produce estimators which cannot be analysed reliably by asymptotic methods". This suggested a minimum sample size of, at least, 60 respondents for each UK and US DCE. Similarly, Lancasar et al.[165] suggest that 30 respondents per block version of the DCE questionnaire should be adequate though *post hoc* analyses will require more. On the other hand, applying Orme's[166] formula meant, for example, a minimum of 84 respondents would be needed for the UK target population.

There are, of course, more statistically sound ways of determining the sample size for the DCEs but these methods either (1) assume each choice task is independent of all other

Figure 5.2: Example choice set for the UK DCE



This person prefers drug B. This means he or she thought it better to have their medicines given to them at home under the supervision of a health worker using a method that doesn't involve needles – even though this means taking their medicines more frequently. They do not seem to be concerned about disruption to their daily activities and the moderate risk of adverse events. This is worth the cost to them. Your own preferences, of course, may be completely different from this.

choice tasks or (2) require some prior knowledge of $\boldsymbol{\beta}$ to calculate either expected choice probabilities or expected standard errors – see Rose and Bleimer[167]. I did not consider these statistical approaches here but determined (given the resources available for this research) that, for each UK and US survey, a sample of, at least, 200 respondents (equivalent to 100 responses for each block version of the questionnaire) should be adequate for meaningful statistical analyses. This sample size is consistent with recommendations made by the ISPOR Good Research Practices for Conjoint Analysis Task Force[168].

5.3.5 Econometric modelling

On completion of the DCE survey, I checked the data collected for incomplete sequence(s) of choices so as to avoid estimation biases due to discontinuous preferences ("noise") created by information overload, boredom, unfamiliarity with or lack of interest in the survey. For the UK DCE, I found that each survey respondent completed all 12 choice tasks. The UK dataset thus provided 10608 usable choice responses from 442 respondents. I make no argument that this sample is representative of the UK population. A summary of the demographic characteristics of the sample is supplied in Appendix 5A. For the US DCE, I also found that each survey respondent completed the sequence of 12 choices. However, for three respondents, there were missing data on some of their characteristics whilst for one respondent I had no information on *all* characteristics. The US dataset thus provided 5040 usable choice responses from 210 survey respondents. As before, I make no claim that this sample is representative of all healthcare professionals in the US. Demographic characteristics of the estimation sample are shown in Appendix 5B. For the UK and US surveys, I had no information on the number of people the vendor approached in order to achieve the minimum number of respondents. It is not possible therefore to compute response rates for the surveys although the choice data was collected from more than the minimum number of respondents specified.

To explore plausible explanations (unobserved heterogeneity, variation in preferences, respondent fatigue etc.) for the observed sequence of choices in the UK and US datasets, I estimated a number of econometric models. As my starting point, I estimated a "mother" MNL model. The IID/IIA assumption underlying this model (with normalization of the scale parameter to one), however, is equivalent to saying that all survey respondents have the same preferences and/or that unobserved variation around these preferences are similar. For this reason, some researchers will argue that all estimates derived from the "mother" MNL model

are biased. I therefore considered alternative econometric models that relax the IID/IIA restriction.

I first considered a heteroskedastic multinomial (HMNL) model, where the scale parameter is no longer normalized to one but considered a variable that must be estimated. The error terms are therefore no longer IID distributed, and a typical approach is to express the scale parameter as a function of a vector of respondents' characteristics (**Z**). The probability of an individual choosing alternative good *j* from among a set of competing goods, in a given choice situation, is then given by:

$$\mathrm{P_{jns}}(\mathrm{y_j}=1)=\frac{\exp(\exp(\alpha\mathbf{Z_s})\,\beta_{jk}\mathbf{X_{jk}})}{\sum_{j=1}^{J}\exp(\exp(\alpha\mathbf{Z_s})\beta_{jk}\mathbf{X_{jk}})} \tag{17}$$

where α is a vector of coefficients reflecting the influence of respondents' characteristics on the error variance. If $\hat{\alpha}$ is not statistically different from zero, we revert back to or close to the "mother" MNL model. If $\hat{\alpha}$ is statistically-significant different from zero, then it is possible to *exogenously* determine subpopulations with somewhat identical preferences[169].

A variant of the HMNL model is the entropy multinomial (EMNL) model in which the scale parameter is a function of entropy (*E*): a measure of the information content or uncertainty represented in the probability distribution of a discrete random variable, in this case the choice variable *y*. In DCE literature, entropy summarizes the impact of task complexity or respondent fatigue due to the number of choice alternatives; the number and correlation between attributes and attribute-levels; and similarity between the alternatives. The relationship between the scale parameter and entropy of each choice situation can be expressed as:

$$\mu_{ns}=\exp(\theta_1\mathrm{E_{ns}}+\theta_2\mathrm{E_{ns}^2}) \tag{18}$$

$$\mathrm{E_{ns}}=-\sum_{j=1}^{J}\widehat{\mathrm{P_{jns}}}\log\left(\widehat{\mathrm{P_{jns}}}\right)$$

where $\widehat{P_{jns}}$ is the estimated choice probability from the "mother" MNL model; θ_1, θ_2 are parameters associated with entropy. The linear term $\widehat{\theta_1}$ measures deviation from maximum entropy, i.e., completely random choices; whilst the quadratic term $\widehat{\theta_2}$ identifies non-linearity in the relationship above. The case of $\widehat{\theta_1} > 0$ indicates entropy is either offset by exertion of more effort and/or (independent of effort) respondents' under-estimation of the differences

between choice alternatives. Researchers often treat the case of $\widehat{\theta_1} < 0$ and $\widehat{\theta_2} > 0$ as indicative of respondent fatigue (declining effort) as a survey respondent works through a sequence of choice sets[170],[171].

I also considered the mixed multinomial (MMNL) model in which $\boldsymbol{\beta}$ varies randomly across individual respondents. Typically, these random coefficients are drawn from a mixture of continuous parametric distributions denoted by $f(\beta|\delta)$, where δ refers to parameters of that mixture distribution. Here I assumed the individual-specific non-price coefficients, that had no statistically-significant effects in the "mother" MNL model, are normally distributed and correlated with a price coefficient that is log-normal distributed and constrained to be negative. This combination of random attribute-coefficients and extreme-value (Gumbel) distributed error terms, however, means that the MMNL model cannot be solved analytically, but approximated via Monte Carlo simulations. The choice probability for alternative good *j* (out of all *J* goods) is then given by:

$$\mathrm{P_{jns}}(\mathrm{y_j} = 1) = \int_{\mathrm{r=1}}^{\mathrm{R}} \left[\frac{\exp(\beta_{\mathrm{sjk}} \mathbf{X}_{\mathrm{jk}})}{\sum_{\mathrm{j=1}}^{\mathrm{J}} \exp(\beta_{\mathrm{sjk}} \mathbf{X}_{\mathrm{jk}})} \right] . \mathrm{f}(\beta|\delta)\, \partial \beta_{\mathrm{sjk}} \tag{19}$$

where *R* is the number of Halton replications; and $\beta_{sjk}$is the r^{th} draw from $f(\beta|\delta)$[172].

Finally, I considered a latent-class multinomial (LCMNL) model that assumes attribute-coefficients are drawn from a mixture of non-parametric discrete distributions, representing *C* latent classes of homogenous subpopulations. It is not known *a priori* which latent class an individual belongs to – so rather than assume fixed class membership, the probability of class membership (π) can be estimated as:

$$\pi_{\mathrm{cs}}(\gamma) = \frac{\exp(\nu_c + \gamma_{\mathrm{c}} \mathbf{Z}_{\mathrm{s}})}{1 + \sum_{\mathrm{c=1}}^{\mathrm{C-1}} \exp(\nu_c + \gamma_{\mathrm{c}} \mathbf{Z}_{\mathrm{s}})} \tag{20}$$

where $\sum_{c=1}^{C} \pi_c = 1$, the vector γ $(= \gamma_1, \gamma_2, \dots, \gamma_C)$ refers to the effect of individuals' characteristics on class membership, and ν_c is a vector of class-specific constants[173]. Unconditional on class membership, the probability (P^*) of observing the sequence of *N* choices that an individual respondent makes is given by:

$$\mathrm{P^*(y_{Ns} = 1)} = \sum_{\mathrm{c=1}}^{\mathrm{C^*}} \pi_{\mathrm{cs}} \prod_{\mathrm{n=1}}^{\mathrm{N}} \prod_{\mathrm{j=1}}^{\mathrm{J}} \left(\frac{\exp(\beta_{\mathrm{cjk}} \mathbf{X}_{\mathrm{jk}})}{\sum_{\mathrm{j=1}}^{\mathrm{J}} \exp(\beta_{\mathrm{cjk}} \mathbf{X}_{\mathrm{jk}})} \right)^{\mathrm{y_s}}$$

where the "optimal" number of latent classes C^* is determined by: (1) estimating a series of LCMNL models with different numbers of latent classes; (2) choosing the preferred model using the lowest consistent Akaike Information Criterion [cAIC] and/or Bayes Information Criterion [BIC]; and (3) making judgements on the trade-off between improved log-likelihoods and increase in standard errors (loss in precision) of the attribute-coefficients as the number of latent-classes gets large.

I estimated the models in STATA v. 11 using the attribute-based (**X**) and respondent-characteristics (**Z**) variables in Tables 5.3 and 5.4 below. The respondent-characteristics variables are those in the shaded regions of Tables 5.3 and 5.4. Following Bech and Gyrd-Hansen[174], the values of the explanatory variables reflect "effects coding" rather than 0-1 dummies.

Table 5.3: Explanatory variables for the UK sample

Variables	Definitions (Effects coding)
INTRAVENOUS	= 1 if a drug is administered intravenously (1, 0, 0, -1)
SUBCUTANEOUS	= 1 if a drug is administered subcutaneously (0, 1, 0, -1)
INTRAMUSCULAR	= 1 if a drug is administered intramuscularly (0, 0, 1, -1). The reference category (-1) is administration via needle-free routes
DOSFREQ	Continuous variable referring to the number of unit administrations over a one year period.
NONCLINICAL_SELF	= 1 if a drug is self-administered in non-clinical settings (1, 0, -1)
NONCLINICAL_SUPV	= 1 if a drug is administered in non-clinical settings under the supervision of a qualified healthcare professional (0, 1, -1). The reference category (-1) is drug administration in clinical settings
DDA_MODERATE1	= 1 if a given mode of administration is associated with moderate but manageable disruption to respondents' daily activities (1, 0, 0, -1)
DDA_MODERATE2	= 1 if a given mode of drug administration is associated with moderate disruptions to daily activities that a respondent cannot cope with (0, 1, 0, -1)
DDA_SEVERE	= 1 if a given mode of drug administration is associated with severe disruption to the respondent's daily activities (0, 0, 1, -1). The reference category (-1) is a mode of administration that carries no risk of disruption to patients' daily activities
RAE_MODERATE	= 1 if the risk of adverse events associated with a given mode of drug administration is moderate (1, 0, -1)
RAE_SEVERE	= 1 if the risk of adverse events associated with a given mode of drug administration is severe (0, 1, -1). The reference category (-1) is drug delivery that is associated with no risk of adverse events
PRICE (COST)	Continuous variable indicating the time and travel costs borne by patients per unit administration
A	Alternative-specific constant = 1 for drug option *A* (1, -1). The reference point, drug option *B* = -1
FEMALE	= 1 if survey respondent is female (1, -1). The reference category (-1) are males
RESPONDENTAGE	Continuous variable indicating the age of a survey respondent
VOCATIONAL	= 1 if the highest level of education attained by a respondent is vocational training (1, 0, -1)
GCSEs_O+A	= 1 if the highest level of education attained by a respondent is GCSEs O' and A' levels (0, 1, -1). The reference category (-1) are respondents with "higher education"
EMPLOYED	= 1 if respondent is employed (1, -1). The reference category (-1) are those who are currently unemployed
INCOME_1	= 1 if respondent's annual household income is under £15,000 (1, 0, 0, 0, -1)
INCOME_2	= 1 if respondent's annual household income is between £15,000 and £29,999 (0, 1, 0, 0, -1)
INCOME_3	= 1 if respondent's annual household income is between £30,000 and £49,999 (0, 0, 1, 0, -1)
INCOME_4	= 1 if respondent's annual household income is between £50,000 and £75,000 (0, 0, 0, 1, -1). The reference category (-1) are respondents with annual household income in excess of £75,000
PRIOR_ ILLNESS	= 1 if a survey respondent had received medical treatment under the advice or guidance of a qualified health worker over the past year (1, -1). The reference category (-1) are those who remained healthy over the past year

Table 5.4: Explanatory variables for the US sample

Variables	Definitions (Effects coding)
INTRAVENOUS	= 1 if a drug is administered intravenously (1,0, 0, -1)
SUBCUTANEOUS	= 1 if a drug is administered subcutaneously (0, 1, 0, -1)
INTRAMUSCULAR	= 1 if a drug is administered intramuscularly (0, 0, 1, -1). The reference category (-1) is administration via needle-free routes
DOSFREQ	Continuous variable referring to the number of unit administrations over a one year period.
NONCLINICAL_SELF	= 1 if a drug is self-administered in non-clinical settings (1, 0, -1)
NONCLINICAL_SUPV	= 1 if a drug is administered in non-clinical settings under the supervision of a qualified healthcare professional (0, 1, -1). The reference category (-1) is drug administration in clinical settings
NONCOMP_MODERATE	= 1 if the risk of patient non-compliance associated with a given mode of drug administration is moderate (1, 0, -1)
NONCOMP_SEVERE	= 1 if the risk of patient non-compliance associated with a given mode or method of drug administration is severe (0, 1, -1). The reference category (-1) is drug delivery that is associated with no risk of non-compliance
RME_MODERATE	= 1 if the risk of medication errors made by health staff or by patients self-administering is moderate (1, 0, -1)
RME_SEVERE	= 1 if the risk of medication errors made by health staff or by patients self-administering is severe (0, 1, -1). The reference category (-1) is drug delivery that is associated with no risk of medication errors
PRICE (COST)	This refers to the cost of resources expended on drug administration per patient per single full treatment course over a year. It excludes drug acquisition costs
A	Alternative-specific constant (intercept) = 1 for the drug *A* option. The reference point, drug option *B* = -1
FEMALE	= 1 if survey respondent is female. The reference category (-1) are males
RESPONDENTAGE	Continuous variable indicating the age of a survey respondent
SOLO	= 1 if a respondent works in a solo medical practice (1, 0, 0, 0, -1)
GOVERNMENT	= 1 if a respondent works in a government-run healthcare institution (0, 1, 0, 0 , -1)
PRIVATE-FOR-PROFIT	= 1 if a respondent works in a private-for-profit healthcare institution (0, 0, 1, 0, -1)
NON-FOR-PROFIT	= 1 if a respondent works in a non-for-profit healthcare institution (0, 0, 0, 1, -1). The reference category (-1) are respondents who work in all other healthcare institutions not mentioned above
INPATIENTS	= 1 if a respondent caters for inpatients (1, 0, 0, -1)
OUTPATIENTS	= 1 if a respondent caters for out-patients (0, 1, 0, -1)
EMERGENCY	= 1 if a respondent caters for patients in accident and emergency departments (0, 0, 1, -1). The reference category (-1) are respondents who cater for all other kinds of patients
ENC	= 1 if a respondent's healthcare institution is located in the East North Central census region (1, 0, 0, 0, 0, 0, 0, 0, -1)
ESC	= 1 if a respondent's healthcare institution is located in the East South Central census region (0, 1, 0, 0, 0, 0, 0, 0, -1)
MA	= 1 if a respondent's healthcare institution is located in the Middle Atlantic census region (0, 0, 1, 0, 0, 0, 0, 0, -1)
Mo	= 1 if a respondent's healthcare institution is located in the Mountain census region (0, 0, 0, 1, 0, 0, 0, 0, -1)

NE	= 1 if a respondent's healthcare institution is located in the New England census region (0, 0, 0, 0, 1, 0, 0, 0, -1)
PA	= 1 if a respondent's healthcare institution is located in the Pacific census region (0, 0, 0, 0, 0, 1, 0, 0, -1)
WNC	= 1 if a respondent's healthcare institution is located in the West North Central census region (0, 0, 0, 0, 0, 0, 1, 0, -1)
WSC	= 1 if a respondent's healthcare institution is located in the West South Central census region (0, 0, 0, 0, 0, 0, 0, 1, -1). The reference category (-1) are respondents who work in healthcare institutions located in the South Atlantic census region

Notes: I control for the location of respondents' healthcare institution to address the possibility that the level and rise in healthcare expenditures (and for that matter resources spent on drug administration) may differ across US census regions.

5.3.6 Measuring patient benefits

The outputs of the models above allow us to compute, first, the marginal willingness-to-pay (MWTP) for a given change in an attribute (level) or a bundle of attributes. MWTP is the marginal rate of substitution between the non-monetary attributes (singly or in a bundle) and the price/cost attribute – assuming there is only one good available that will be chosen with a 100% certainty. The $\widehat{\text{MWTP}}$ for a *single* non-monetary attribute = $-\widehat{\beta_k}/\widehat{\beta_p}$. Classical confidence intervals were generated using 100 bootstrap replicates of $\widehat{\text{MWTP}}$. Admittedly, a higher number of replicates is needed for more precise estimation but I prefer this procedure as: (1) it is computationally less demanding; (2) it uses actual data from respondents without making parametric assumptions about the distribution of $\widehat{\text{MWTP}}$[175]; and (3) it is compatible with all STATA estimators for the HMNL, EMNL, MMNL and LCMNL models.

Second, I estimated the incremental welfare gain or loss from switching (changing) from one good to another using the expected compensating variation (ECV). This is a more valid measure of welfare benefits when there is uncertainty about which good will be chosen. For discrete-choice probabilities estimated using an MNL-type model, ECV is formally computed as follows:

$$\widehat{\text{ECV}} = -\frac{1}{\hat{\lambda}}\left[\ln\sum_{j=1}^{J}\exp(\widehat{V_j^0}) - \ln\sum_{j=1}^{J}\exp(\widehat{V_j^1})\right] \tag{22}$$

where λ is the marginal utility of income proxied by the negative coefficient of the price/cost attribute; and the superscripts 0 and 1 denote the conditions before and after the change (switch). The log-sum expressions or "inclusive values" in the brackets effectively weight the systematic utilities by the probability that an alternative good will be chosen in each state. Analogous to a change in consumer surplus, ECV measures the amount of money that will have to be extracted from an individual for them to remain indifferent between the initial (0) and final (1) states[176],[177],[178].

For an example of valuing "product innovations" using this ECV metric, see Trajtenberg's [179] work on computed axial tomography scanners. Here the "product innovation" is different modes of drug administration – and the reported ECV estimates provide a single-period monetary value of what might be considered "intangible benefits" of giving end-users what they want in terms of the mode of drug delivery. I did not consider intertemporal flows of benefits: my ECV estimates are not expressed in discounted present

values. But from Chapter 4 and previous work[139], we know there are potential savings in administration costs from reformulating or reverse-engineering a drug product. To ascertain whether these "intangible" benefits are of any significant importance, I compared the ECV estimates with predicted administration cost savings of switching from one mode of drug delivery to another.

5.4. Results

5.4.1 Sample UK general public

5.4.1.1 Choice probability

Table 5.5 shows the results from the econometric modelling. Given that I employed a forced-choice format for the survey with only two alternatives, I could not perform a Hausman-McFadden[180] statistical test for the IID/IIA assumption underlying the MNL model. However, the different results obtained from HMNL, EMNL[15], MMNL and LCMNL models suggest that the IID/IIA assumption would have been violated. The HMNL and EMNL models indicate non-constant error terms, whilst the MMNL and LCMNL models indicate variation in observed preferences across the survey respondents. Based on the log-likelihoods and Akaike Information Criterion (AIC), the LCMNL model with two latent classes[16] offers the best model fit to the UK choice data. That is to say, preference variation in the dataset can be conveniently represented by two homogenous subgroups of respondents. The chances that any individual belongs to the first subgroup (latent-class 1) is determined by age, gender, and educational attainment.

That said, the other models provide useful insights on the choice behaviour of respondents. For example, a Lagrangian Multiplier test for heteroskedastic errors in the

[15] Because the measure of entropy is derived from choice probabilities predicted from the MNL model, it is in effect an "endogenous" explanatory variable that might bias the coefficients of the EMNL model. In Table 5.5, one observes that, even if these coefficients are biased, they are consistent with that of the MNL and HMNL models. What is more, the primary purpose of the EMNL model is to assess respondents' (fatigued) reactions to the choice tasks, i.e., we are interested mostly in equation (18) and not equation (17).

[16] In determining C^*, I estimated a number of LCMNL models with 2-10 latent-classes. To increase the speed of computation, I estimated these models without the **Z** variables in the class-membership function, i.e., equation (20). I found the closest competitor to the LCMNL with two-classes (cAIC = 5522.461, BIC = 5495.461) was an LCMNL with four classes (cAIC = 5528.196, BIC = 5473.196). I chose the two-class LCMNL model on the basis of precision loss in the estimated coefficients: the fourth latent-class of the LCMNL model with four-classes had no statistically-significant coefficients.

HMNL model showed statistically-significant unobserved variation that is explained mostly by gender and education status[17]. This is not surprising as these respondent characteristics are well-known determinants of how (imperfectly) informed people are. Likewise, a Lagrangian Multiplier test for heteroskedastic errors in the EMNL model showed statistically-significant unobserved variation. However, the statistically-insignificant entropy parameters ($\widehat{\theta_1}, \widehat{\theta_2}$) of the EMNL model, with $\widehat{\theta_1} > 0$, suggested that respondent fatigue (perhaps offset by "learning effects") could be ignored as plausible explanations for the sequence of choices observed.

Focusing on the results of the LCMNL model, I observed that, conditional on membership of latent-class 1, the average or representative survey respondent is indifferent to needle-free modes of drug administration when compared with intravenous or subcutaneous routes conditional on the other attributes. Respondents are indifferent in the sense that the coefficients for variables for intravenous and subcutaneous modes of administration were not statistically different from zero. I interpret this to mean respondents are informed enough to know that, in some disease states, needle-free routes may not be the best or a feasible method of drug administration. On the other hand, respondents, on average, show a negative preference for intramuscular modes of drug delivery when compared with needle-free routes, perhaps because of the pain involved. Similarly, I observed a negative preference for drug administration modes that involve higher dosing frequency albeit the magnitude of the effect was "small" and close to zero. I observed also a positive preference for self-administration within a non-clinical setting and a negative statistically-insignificant preference for drug administration in non-clinical settings under supervision. A probable explanation for this result is that if administering a drug requires supervision by a qualified healthcare professional, then one might be better off having the drug administered in a clinical setting.

[17] Not reported in Table 5.5, for brevity reasons, is that coefficient for age = 0.010 (p value < 0.001) and that for employment status = -0.091 (p value < 0.05) are statistically-significant contributors to unobserved heterogeneity in choices. Other respondents' characteristics: prior illness within the past year and household-income were only statistically-significant at the 10% level.

Table 5.5: Econometric results

Dependent variable: CHOICE PROBABILITY						
	MNL model	HMNL model	EMNL model	MMNL model	LCMNL model	
Variables/ Coefficients:	$\hat{\beta}$(SE)	$\hat{\beta}$(SE)	$\hat{\beta}$(SE)	$\widehat{\beta_s}$(SE)	$\hat{\beta}_1$(SE)	$\hat{\beta}_2$(SE)
INTRAVENOUS	-0.011 (0.091)	0.005 (0.057)	0.017 (0.042)	-0.026 (0.110)	-0.119 (0.439)	0.021 (0.123)
SUBCUTANEOUS	0.021 (0.043)	0.016 (0.027)	0.002 (0.019)	-0.016 (0.049)	-0.073 (0.130)	0.070 (0.060)
INTRAMUSCULAR	-0.136 (0.049)**	-0.090 (0.031)**	-0.111 (0.037)**	-0.123 (0.053)*	-0.560 (0.210)**	-0.219 (0.064)***
DOSFREQ	-0.001 (0.000)***	-0.0005 (0.000)***	-0.001 (0.000)**	-0.001 (0.000)***	-0.003 (0.001)***	-0.001 (0.000)**
NONCLINICAL_SELF	0.167 (0.037)***	0.108 (0.025)***	0.112 (0.038)**	0.210 (0.039)***	0.424 (0.128)***	0.146 (0.045)**
NONCLINICAL_SUPV	-0.122 (0.028)***	-0.078 (0.019)***	-0.067 (0.025)**	-0.120 (0.030)***	-0.124 (0.093)	-0.098 (0.037)**
DDA_MODERATE1	0.391 (0.033)***	0.239 (0.033)***	0.214 (0.074)**	0.437 (0.038)***	1.210 (0.150)***	0.264 (0.050)***
DDA_MODERATE2	-0.267 (0.042)***	-0.171 (0.032)***	-0.167 (0.059)**	-0.300 (0.046)***	-1.058 (0.141)***	-0.075 (0.061)
DDA_SEVERE	-0.525 (0.039)***	-0.336 (0.044)***	-0.308 (0.103)**	-0.588 (0.044)***	-1.376 (0.184)***	-0.324 (0.057)***
RAE_MODERATE	0.169 (0.034)***	0.108 (0.024)***	0.111 (0.041)**	0.212 (0.038)***	0.522 (0.141)***	0.067 (0.043)
RAE_SEVERE	-0.743 (0.039)***	-0.461 (0.056)***	-0.406 (0.138)**	-0.869 (0.044)***	-2.342 (0.215)***	-0.204 (0.059)***
PRICE (COST)	-0.008 (0.001)***	-0.005 (0.001)***	-0.004 (0.001)**	-0.0118 (0.040)***	-0.012 (0.002)***	-0.008 (0.001)***
A	-0.054 (0.054)	-0.033 (0.034)	0.002 (0.025)	-0.037 (0.066)	0.370 (0.274)	-0.122 (0.073)!
Entropy($\widehat{\theta_1}, \widehat{\theta_2}$)	—	—	(1.026, 0.717)	—	—	
$\hat{\alpha}_0$(FEMALE)	—	0.170 (0.035)***	—	—	—	
$\hat{\alpha}_1$(VOCATIONAL)	—	-0.230 (0.085)**	—	—	—	
$\hat{\alpha}_2$(GCSEs_O+A)	—	-0.199 (0.054)***	—	—	—	
$\hat{\pi}_c$	—	—	—	—	0.492	0.508
$\hat{\gamma}_c$(VOCATIONAL)	—	—	—	—	-0.892 (0.324)**	—
$\hat{\gamma}_c$(GCSEs_O+A)	—	—	—	—	0.577 (0.203)**	—
$\hat{\gamma}_c$(FEMALE)	—	—	—	—	0.285 (0.122)*	—
$\hat{\gamma}_c$(RESPONDENTAGE)					0.024 (0.009)**	
Model statistics						
Log-likelihood (AIC)	-2855.3566 (5736.713)	-2812.1903 (5670.381)	-2837.9705 (5670.381)	-2782.5096 (5611.019)	-2652.879 (5379.757)	

Notes: SE = standard error. For the HMNL and LCMNL models, I report *selected* effects of individuals' characteristics on the scale-parameter and latent-class membership. MMNL model was estimated using 500 Halton draws with correlated normally-distributed coefficients for the variables A, INTRAVENOUS and SUBCUTANEOUS and a log-normal distribution for the price coefficient. *** p < 0.001 ** p < 0.01 * p < 0.05 ! p < 0.10. AIC = Akaike Information Criterion.

The results show a positive preference for “moderate but manageable” disruptions to daily activities and a negative preference for “severe” or “moderate but unmanageable” disruptions to daily activities. This suggests that respondents did take into account the hypothetical nature of the choice tasks: although possible (in the future), a mode of drug administration that is associated with zero disruption to daily activities may not be currently available or technologically feasible. Similarly, I observed a positive preference for modes of drug administration associated with a “moderate” risk of adverse events and a negative preference for modes of drug administration associated with “severe” risk of adverse events. Again, I observed, on average, some kind of mental accounting of the fact that a mode of drug delivery that has a zero risk of adverse events may not be available or technologically feasible. For individuals belonging to latent-class 1, the coefficient for the alternative-specific-constant *A* is not statistically significant albeit the magnitude of the coefficient suggests a skewed preference for drug option *A* (holding all other attributes and attribute-levels the same).

Conditional on membership of latent-class 2, I observed similar choice patterns with the following exceptions: the coefficients of the variables for moderate disruptions-to-daily-activities and moderate risks of adverse events are not statistically significant. That respondents belonging to latent-class 2 are indifferent to the reference category for these attribute-levels, provides further support to the argument that respondents may have, in their decision choices, considered that drug delivery modes with zero disruptions to their daily activities and/or zero risk of adverse events are perhaps unavailable even though they are desirable. The magnitudes of the coefficients of the variables for intravenous and subcutaneous administration indicate a positive preference for these modes of drug administration relative to needle-free routes of administration. But since these effects are not statistically different from zero, I maintain the argument that respondents are generally indifferent to the choice between needle-free and intravenous/subcutaneous routes of drug administration. Also the negative sign for the alternative-specific constant *A* suggests a tendency to choose alternative drug option *B* (holding all other attributes and attribute-levels the same). The effect is, however, only significant at the 10% level.

For both latent-classes, I observed a small but statistically-significant coefficient for the price/ cost attribute. This suggests that UK survey respondents have price inelastic “demands” for the attributes of drug administration investigated in response to any (out-of-pocket) costs of accessing healthcare. This probably reflects two things. One, the fact that the

UK NHS provides tax-funded insurance protection against the financial risks of ill health; and two, that the costs in question are by and large 'unavoidable': without spending resources on some form of a vehicle for administering a drug, patients will be unable to realize the (incremental) health benefits offered by that drug.

5.4.1.2 Patient benefits

Table 5.6 shows estimates of the MWTP for each of the non-price attributes studied and the associated confidence intervals around these estimates. A positive $\widehat{\text{MWTP}}$ indicates a preference for an attribute taking into account the associated price/cost, whilst a negative $\widehat{\text{MWTP}}$ indicates a dispreference for an attribute net of the associated price/cost. As observed in Table 5.6, there are subtle differences in the MWTP estimates obtained from the different econometric models. I focussed on estimates from the LCMNL model as this provided the best fit with the choice data collected. This shows a statistically significant and substantial willingness-to-pay for drug delivery modes that are associated with "moderate" disruption to daily activities and "moderate" risk of adverse events. There is also a statistically-significant and substantial willingness-to-pay to avoid drug delivery modes that are associated with "moderate but unmanageable" or "severe" disruptions to daily activities. Similarly, there is a statistically significant and substantial willingness-to-pay to avoid drug delivery modes that are associated with "moderate" risk of adverse events. Further, there is a statistically-significant and substantial willingness-to-pay to avoid the intramuscular route.

To evaluate the welfare change, i.e., the "intangible benefits" from manufacturing drugs in a patient-friendly manner, I considered the following. A given biologic drug *C* can be manufactured in two ways (C1 and C2). Assume, as in the survey, that both versions of the drug have the same molecule, efficacy and safety profile. In option C1, the drug can be manufactured for intravenous administration in clinical settings, and this mode of drug delivery is associated with "severe" risk of adverse events and "severe" disruptions to patients' daily activities. Option C2 is where a drug is manufactured for subcutaneous self-administration in non-clinical settings and this mode of drug delivery is associated with "moderate" risk of adverse events and "moderate" disruptions to patients' daily activities. In this case, one can compute the expected compensating variation ($\widehat{\text{ECV}}$) using equation (22). Since household-income categories had no statistically-significant effect on class membership in the preferred LCMNL model, I do not differentiate $\widehat{\text{ECV}}$ by household-income category.

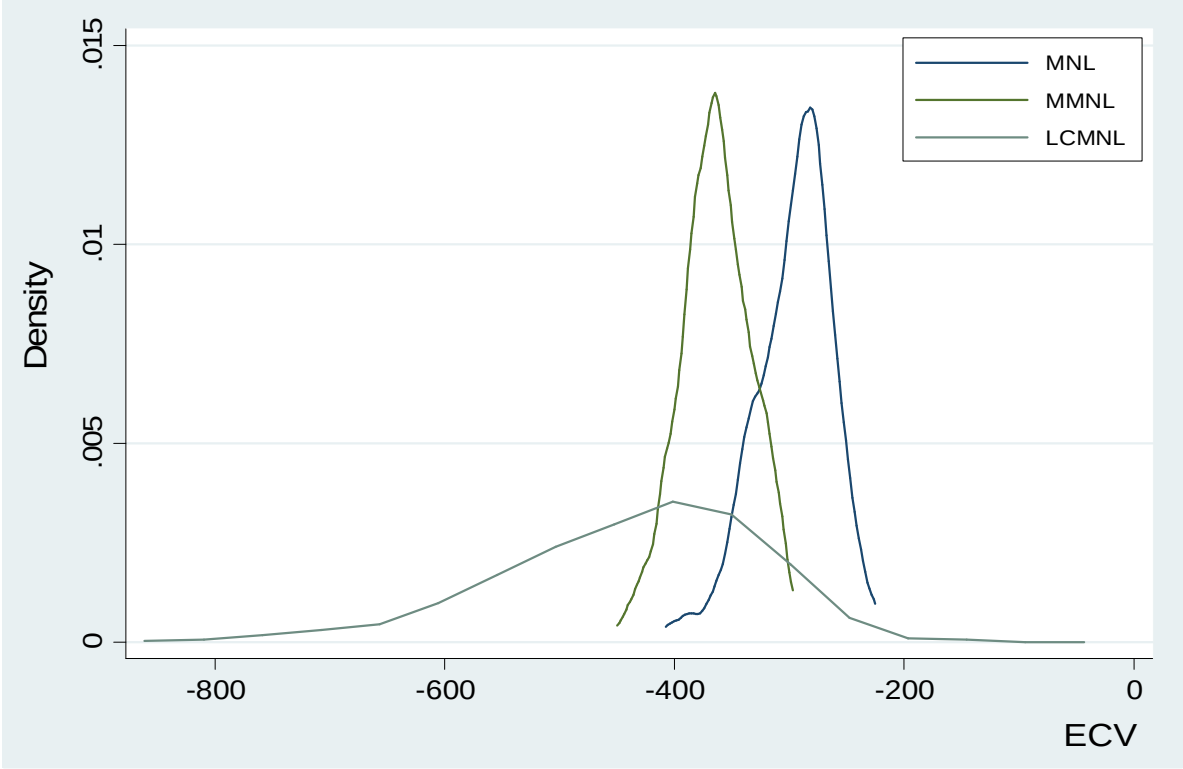
Table 5.6: Marginal willingness-to-pay estimates

	MNL model	HMNL model	EMNL model	MMNL model	LCMNL model
Variables:	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)
INTRAVENOUS	-2.123 (-4.089, -0.156)	1.169 (-0.742, 3.080)	3.433 (1.726, 5.139)	2.311 (-2.864, 7.485)	5.848 (-50.374, 62.070)
SUBCUTANEOUS	2.347 (1.374, 3.320)	2.600 (1.576, 3.625)	0.226 (-0.516, 1.027)	-8.325 (-10.307, -6.343)	2.417 (-7.759, 12.593)
INTRAMUSCULAR	-17.433 (-18.645, -16.222)	-19.192 (-20.413, -17.971)	-26.079 (-27.152, -25.006)	-9.385 (-10.997, -7.773)	-40.634 (-72.472, -8.796)
DOSFREQ	-0.082 (-0.086, -0.078)	-0.090 (-0.094, -0.085)	-0.123 (-0.127, -0.119)	-0.098 (-0.104, -0.092)	-0.234 (-0.296, -0.172)
NONCLINICAL_SELF	21.448 (20.508, 22.389)	22.412 (21.478, 23.345)	26.354 (25.548, 27.160)	31.005 (29.869, 32.141)	28.569 (21.401, 35.738)
NONCLINICAL_SUPV	-15.446 (-16.061, -14.830)	-15.738 (-16.362, -15.114)	-15.456 (-15.950, -14.961)	-19.828 (-20.639, -19.016)	-15.532 (-26.038, -5.025)
DDA_MODERATE1	50.090 (48.912, 51.267)	49.397 (48.147, 50.647)	50.144 (49.015, 51.273)	58.883 (57.652, 60.114)	84.705 (67.590, 101.819)
DDA_MODERATE2	-33.997 9-35.148, -32.847)	-35.239 (-36.507, -33.970)	-38.860 (-39.891, -37.829)	-43.131 (-44.347, -41.915)	-58.841 (-72.278, -45.404)
DDA_SEVERE	-66.363 (-67.738, -64.987)	-68.641 (-70.141, -67.141)	-71.549 (-72.795, -70.303)	-73.642 (-75.149, -72.136)	-94.502 (-117.906, -71.097)
RAE_MODERATE	22.060 (21.196, 22.923)	22.636 (21.744, 23.529)	26.368 (25.594, 27.142)	30.318 (29.226, 31.410)	23.492 (17.286, 29.699)
RAE_SEVERE	-66.363 (-67.738, -64.987)	-68.641 (-70.141, 67.141)	-71.549 (-72.795, -70.303)	-73.642 (-75.149, -72.136)	-94.502 (-117.906, -71.097)

Notes: The 95% CIs above are "standard or classical confidence intervals" calculated using 100 bootstrapped replicates of MWTP. This is because accurate, less-erratic and reliable "bootstrap confidence intervals" require replications in the order of 1000, which would have been computationally demanding and time consuming [181]. The confidence intervals reported are therefore not exact.

Based on the MNL model, switching from option C1 to C2 yields a welfare gain ($\widehat{ECV}$ per patient per unit administration) of -£296 (95 % CI: -£302 to -£289). Based on the MMNL model, $\widehat{ECV}$ is: -£364 (95% CI: -£370 to -£358)[18]. Based on the LCMNL model with two latent-classes, and unconditional on class membership, I obtained $\widehat{ECV}$ of -£435 (95% CI: -£524 to -£349). The negative sign reflects the fact that ECV is the amount of money that has to be taken from the state of having option C2 so that the average respondent will be indifferent to option C1. Also, with the UK choice data, failing to control for preference heterogeneity in the MNL model leads to an underestimation of welfare change. The difference in $\widehat{ECV}$ from the MMNL and LCMNL models most likely reflects differences in the assumed distributions for preference heterogeneity.

Figure 5.3: Sample distribution of ECV estimates



Notes: $\widehat{ECV}$ refers to the welfare change of switching from drug option C1 to C2. The plots are based on 100 bootstrapped estimates of ECV. Large positive and negative $\widehat{ECV}$ outliers from the LCMNL model have been trimmed out to visually enhance comparability with the other models.

[18] The price/cost coefficient reported in Table 5.5 for the MMNL model ($\widehat{\beta_p}$) was derived from $\ln(\widehat{\beta_p})$ using the following retransformation: $\widehat{\beta_p} = \exp(\ln(\widehat{\beta_p}) + 0.5SE^2)$. This, however, assumes the individual $\ln(\widehat{\beta_p})$ recovered from the UK dataset are normally-distributed with a constant variance. If this assumption doesn't hold, one obtains a biased and less consistent retransformed price coefficient, as reported in Table 5.5. Therefore, in computing C1→C2 welfare change, I used the price coefficient from an unreported MMNL model (log-likelihood = -2827.52) with a normally-distributed price coefficient ($\widehat{\beta_p}$= 0.0085) as this figure is comparable with the price coefficients from the other models.

Figure 5.3 shows the kernel density plots of the distribution of $\widehat{\mathrm{ECV}}$ from 100 bootstrap replications using the results from the MNL, MMNL and LCMNL models. The distribution plots for the models overlap. This is reassuring as the models make different assumptions: homogenous preferences in the case of the MNL model and individual preference heterogeneity in the case of the MMNL and LCMNL models. The wider spread of the kernel density plot of $\widehat{\mathrm{ECV}}$ derived from the LCMNL model (relative to the MNL or MMNL models) is perhaps because the LCMNL estimates are based on the average coefficients over the two-latent classes (with weights given by the probability of latent-class membership). Hence if the probabilities of class-membership do not remain constant or fixed for each of the 100 bootstrap samples, one would obtain a lot more variation in $\widehat{\mathrm{ECV}}$ than if (fortuitously) each bootstrap sample has the same class-membership probabilities as reported in Table 5.5. That is, if the proportion of individuals belonging to each of the latent classes does not vary from one bootstrap sample to the other.

5.4.1.3 Patient benefits versus cost savings to the NHS

How do the ECV estimates compare with the direct monetary savings on drug administration costs? To answer this question, one needs to compute savings on the direct administration costs of switching from drug C1 to C2 using the regression-based algorithm described in Chapter 4. See also Tetteh and Morris[139]. I made the following additional assumptions: (1) both drugs C1 and C2 are indicated for management of a chronic illness; (2) product C2 is sold bundled with some of the equipment and consumables used in drug administration; and (3) a single full treatment course of C1 over a year requires 10 unit administrations whilst a single full treatment course of C2 requires 5 unit administrations over a year.

The comparison can be made as follows:

$$\begin{aligned} ln\widehat{ADMINCOST}_{C1} &= 7.1499 - 3.2026(0) - 5.2737(0) + .428(10) + .404(0) - .2896(1) \\ &\quad - .00106(10^2) - .3173(10)(1) \end{aligned}$$

$$\widehat{ADMINCOST}_{C1} = \exp(7.8613) \bullet \hat{\phi}(= 1.0792) = £2800.41$$

$$\widehat{lnADMINCOST_{C2}}$$
$$= 7.1499 - 3.2026(1) - 5.2737(0) + .428(5) + .404(1) - .2896(1)$$
$$- .00106(5^2) - .3173(5)(1)$$

$$\widehat{ADMINCOST_{C2}} = \exp(4.6099) \bullet \hat{\phi}(= 1.3799) = £138.64$$

where $\hat{\phi} = 1.0792$ and 1.3799 are the subgroup-specific smearing factors for intravenous and subcutaneous products respectively.

Assuming the same price per given dose, this yields a cost saving of roughly £2662 per patient per year or an average cost saving of £532 per patient per unit administration of C2 (relative to C1). This is comparable to the absolute value of $\widehat{ECV}$: £435 derived from the LCMNL model.

5.4.2 Sample US healthcare professionals

5.4.2.1 Choice probability

Table 5.7 above shows the results from my econometric analyses. Given the forced choice format with two alternative options, I could not conduct a Hausman-McFadden [180] test for the IID/IIA assumption underlying the MNL model. However, I found that the bias introduced by the IID/IIA assumption is not the great (see below). Based on the log-likelihoods and AIC, the LCMNL model (with $C^* = 2$) offers the best fit to the US choice data. Further support for the LCMNL model is provided by similarity with results for the MMNL model.

Table 5.7: Model outputs

Dependent variable: CHOICE PROBABILITY						
	MNL model	HMNL model	EMNL model	MMNL model	LCMNL model	
Variables/ Coefficients:	$\hat{\beta}$(SE)	$\hat{\beta}$(SE)	$\hat{\beta}$(SE)	$\widehat{\beta_s}$(SE)	$\hat{\beta}_1$(SE)	$\hat{\beta}_2$(SE)
INTRAVENOUS	0.261 (0.291)	0.146 (0.217)	0.129 (0.151)	0.149 (0.351)	0.343 (0.386)	1.612 (1.046)
SUBCUTANEOUS	0.519 (0.103)***	0.401 (0.103)***	0.222 (0.244)	0.453 (0.120)***	0.740 (0.141)***	0.125 (0.354)
INTRAMUSCULAR	-0.111 (0.105)	-0.063 (0.079)	-0.053 (0.059)	-0.200 (0.124)	-0.259 (0.145)!	-0.231 (0.376)
DOSFREQ	-0.003 (0.000)***	-0.002 (0.000)***	-0.001 (0.001)	-0.003 (0.000)***	-0.004 (0.000)***	-0.002 (0.001)
NONCLINICAL_SELF	0.442 (0.056)***	0.322 (0.073)***	0.197 (0.183)	0.481 (0.060)***	0.513 (0.067)***	0.429 (0.189)*
NONCLINICAL_SUPV	-0.097 (0.050)!	-0.059 (0.039)	-0.043 (0.045)	-0.127 (0.056)*	-0.060 (0.063)	-0.439 (0.170)**
NONCOMP_MODERATE	-0.070 (0.050)	-0.044 (0.039)	-0.021 (0.037)	-0.143 (0.064)*	-0.077 (0.067)	-0.111 (0.131)
NONCOMP_SEVERE	-0.200 (0.045)***	-0.149 (0.043)***	-0.083 (0.091)	-0.120 (0.052)*	-0.337 (0.065)***	0.205 (0.143)
RME_MODERATE	0.006 (0.052)	-0.003 (0.040)	0.006 (0.022)	0.042 (0.071)	-0.170 (0.072)*	0.521 (0.143)***
RME_SEVERE	-0.339 (0.041)***	-0.260 (0.061)***	-0.145 (0.155)	-0.445 (0.046)***	-0.075 (0.052)	-1.528 (0.206)***
PRICE (COST)	-0.00006 (0.000)***	-0.00004 (0.000)***	-0.00002 (0.000)	-0.0006 (0.0028)*	-0.00006 (0.000)***	-0.00005 (0.000)***
A	-0.329 (0.092)***	-0.236 (0.080)**	-0.146 (0.145)	-0.323 (0.108)**	-0.418 (0.124)***	-0.696 (0.322)*
Entropy($\widehat{\theta_1}, \widehat{\theta_2}$)	—	—	(3.748, -3.781)	—	—	
$\hat{\alpha}_0$(PRIVATE-FOR-PROFIT)	—	0.210 (0.098)*	—	—	—	
$\hat{\alpha}_1$(NON-FOR-PROFIT)	—	0.269 (0.100)**	—	—	—	
$\hat{\alpha}_2$(INPATIENTS)	—	-0.191 (0.092)*	—	—	—	
$\hat{\alpha}_3$(MA)	—	0.254 (0.125)*	—	—	—	
$\hat{\pi}_c$	—	—	—	—	0.718	0.282
$\hat{\gamma}_c$(INPATIENTS)	—	—	—	—	0.820 (0.360)*	
Model statistics						
Log-likelihood (AIC)	-1446.647 (2917.293)	-1416.264 (2888.527)	-1445.726† (2919.453)	-1399.013 (2852.026)	-1325.1643 (2732.328)	

Notes: SE = standard error. For the HMNL and LCMNL models, I report selected results of the effects of individuals' characteristics on the scale-parameter and latent-class membership. MMNL model was estimated using 500 Halton draws with correlated normally-distributed coefficients for the variables: INTRAVENOUS, INTRAMUSCULAR, NONCOMP_MODERATE and RME_MODERATE; and a log-normal distribution for the price coefficient. *** p < 0.001 ** p < 0.01 * p < 0.05 ! p < 0.10. AIC = Akaike Information Criterion. MA = Middle-Atlantic census region. †Chi-square test for the log-likelihood ratio (for a comparison with the MNL model as the null) failed to reach statistical significance.

That said the HMNL and EMNL models provide additional insights as to the nature of the choices observed. First, note that the coefficients of the HMNL model are quite similar to that of the MNL model, whilst a Lagrangian Multiplier test for heteroskedasticity indicated statistically insignificant unobserved heterogeneity (p value = 0.1064). This preference certainty or minimal residual unobserved variance appears to be explained by whether US healthcare professionals work in private-for-profit or not-for-profit healthcare institutions, and whether they cater for inpatient healthcare demands. It also appears to be explained by whether a given healthcare professional works in the Middle-Atlantic census region as opposed to the South-Atlantic census region (which is the reference category).

Second, note the loss of statistical significance of the attribute-coefficients obtained from the EMNL model. A possible reason is the statistically-insignificant heteroskedasticity identified from the HMNL model. We know from equation (4) that the EMNL model estimates a quadratic relationship between the scale parameter and a measure of entropy (which indicates the degree of randomness or "unlikeliness" of the choice dataset). It appears that, in the presence of statistically-insignificant unobserved heterogeneity, the EMNL model fails to identify appreciable entropy within the US choice data. Consequently, the EMNL model does not significantly fit the data better than the MNL model: a chi-square test for the log-likelihood ratio failed to reach statistical significance (p value = 0.3984) – although all parameters of the EMNL model are identified. In addition, a Lagrangian Multiplier test for error heteroskedasticity in the EMNL model indicated statistically-insignificant unobserved heterogeneity. These findings, plus the statistically insignificant entropy parameters ($\widehat{\theta_1}, \widehat{\theta_2}$), mean we can safely ignore any appreciable respondent-fatigue effects.

Put together, the HMNL and EMNL models suggest almost deterministic (more consistent) choices are made by US healthcare professionals, which is perhaps not surprising considering they are meant to act as informed, near-perfect agents on behalf of their patient populations.

Focusing on the LCMNL model, I observed that, conditional on membership of latent-class 1, the average or representative respondent was indifferent to the choice between intravenous and needle-free modes of drug administration. (Holding all else equal, the coefficient of the variables for intravenous administration was not statistically significant.) The results suggested a positive preference for subcutaneous modes of drug administration and a negative preference for intramuscular drug delivery – albeit the latter is only statistically significant at the 10% level. I observed a negative preference for drug delivery

modes that involve high dosing frequency although the magnitude of this effect is small. Note also the positive preference for self-administration by patients in non-clinical settings. On the other hand, the average respondent was indifferent to drug administration in non-clinical settings under the supervision of a qualified healthcare professional. This offers further support to the notion that if administering a drug in a non-clinical setting requires supervision by a qualified healthcare professional, then one might be better off administering that drug in a clinical setting. The average respondent was willing to accommodate drug delivery methods that are associated with a "moderate" risk of treatment non-compliance (in that the coefficient of the variable for moderate risks of treatment non-compliance was not statistically significant) but exhibited a negative preference for drug delivery modes that carry a "severe" risk of patient non-compliance.

That I observed a negative preference for even "moderate" risks of medication errors possibly reflects concerns about patient safety, professional reputation and medical malpractice suits. Note also the negative coefficient of the variable for severe risks of medication error, although this effect is not statistically significant. As expected, members of latent-class 1 exhibit price-inelastic "demands" for the attributes studied; and the statistically-significant alternative-specific-constant *A* suggests some kind of lexicographic preference for alternative option *B* (listed on the right side of each choice set in the survey) when compared to option *A* (which was listed on the left).

The results for latent-class 2 differ in the following ways: the variables for subcutaneous and intramuscular modes of administration and dosing frequency all have statistically-insignificant effects. The coefficient of the variable for severe risk of treatment non-compliance also had no statistically-significant effects. I observed, however, a statistically-significant negative preference for drug administration in non-clinical settings under the supervision of a qualified healthcare professional. The results indicate a positive (less risk-averse) preference for modes of drug delivery that are associated with "moderate" risk of medication errors but a negative preference for those that carry a "severe" risk of medication errors.

It is worth mentioning that latent-class membership is dictated by whether US healthcare professionals cater for inpatient healthcare demands: the variable INPATIENTS was the only statistically-significant predictor of latent-class membership. Another point to bear in mind is: the MNL model is equivalent to an LCMNL model with one homogenous latent class. Hence, given the LCMNL model with two classes show there is an over 70% probability of a respondent belonging to latent-class 1, we should expect close similarity

between the MNL and LCMNL models. Indeed, the Bayesian Information Criterion for the MNL model (2995.595) is slightly lower than that for the LCMNL model (2999.467) – although the latter offers a substantial improvement in the log-likelihoods. Considering the statistically-insignificant unobserved heterogeneity identified by the HMNL and EMNL models, we can say the US choice data is one instance where estimating an MNL model without consideration of alternative models might not lead to grossly misleading conclusions.

5.4.2.2 End-user benefits

Table 5.8 shows estimates of the MWTP for the non-price attributes studied and the associated 95% confidence intervals. As expected, one observes differences in the sign and magnitude of $\widehat{MWTP}$ from the different econometric models. Focusing on the LCMNL model, these figures indicate a high $\widehat{MWTP}$ for intravenous and subcutaneous modes of drug delivery compared to needle-free routes of drug administration – plus a high $\widehat{MWTP}$ to avoid intramuscular modes of drug administration. Observe also the high $\widehat{MWTP}$ for self-administration in non-clinical settings and a high $\widehat{MWTP}$ to avoid drug delivery modes that are associated with a “severe” risk of treatment non-compliance and/or “severe” risk of medication errors.

Table 5.8: Marginal willingness-to-pay estimates (in $1000)

	MNL model	HMNL model	EMNL model	MMNL model	LCMNL model
Variables:	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$ (95% CI)	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)	$\widehat{MWTP}$(95% CI)
INTRAVENOUS	4.564 (3.688, 5.441)	2.525 (1.580, 3.470)	6.336 (5.037, 7.634)	2.167 (0.954, 3.380)	10.674 (8.533, 12.815)
SUBCUTANEOUS	9.327 (8.923, 9.730)	9.904 (9.552, 10.265)	9.443 (8.975, 9.910)	7.290 (6.833, 7.748)	10.612 (0.713, 11.512)
INTRAMUSCULAR	-1.910 (-2.257, -1.562)	-1.143 (-1.486, -0.081)	-2.492 (-2.990, -1.994)	-4.708 (-5.203, -4.212)	-4.132 (-4.996, -3.269)
DOSFREQ	-0.049 (-0.051, -0.048)	-0.051 (-0.053, -0.501)	-0.051 (-0.052, -0.049)	-0.048 (-0.050, -0.047)	-0.058 (-0.062, -0.055)
NONCLINICAL_SELF	7.815 (7.587, 8.043)	7.802 (7.605, 7.999)	9.015 (8.595, 9.435)	9.075 (8.875, 9.292)	8.211 (7.656, 8.766)
NONCLINICAL_SUPV	-1.597 (-1.761, -1.433)	-1.343 (-1.497, -1.189)	-1.678 (-1.882, -1.473)	-2.700 (-2.878, -2.522)	-1.906 (-2.208, -1.605)
NONCOMP_MODERATE	-1.271 (-1.433, -1.110)	-0.999 (-1.172, -0.825)	-0.664 (-0.886, -0.442)	-3.529 (-3.781, -3.278)	-1.712 (-2.032, -1.394)
NONCOMP_SEVERE	-3.578 (-3.716, -3.439)	-3.526 (-3.659, -3.392)	-3.545 (-3.728, -3.362)	-1.423 (-1.614, -1.232)	-3.741 (-4.039, -3.444)
RME_MODERATE	0.044 (-0.106, 0.193)	-0.207 (-0.380, -0.033)	0.502 (0.274, 0.731)	1.054 (0.665, 1.444)	-0.160 (-0.469, 0.150)
RME_SEVERE	-5.980 (-6.243, -5.717)	-6.218 (-6.453, -5.982)	-6.104 (-6.427, -5.781)	-8.346 (8.597, -8.095)	-7.585 (-8.143, -7.028)

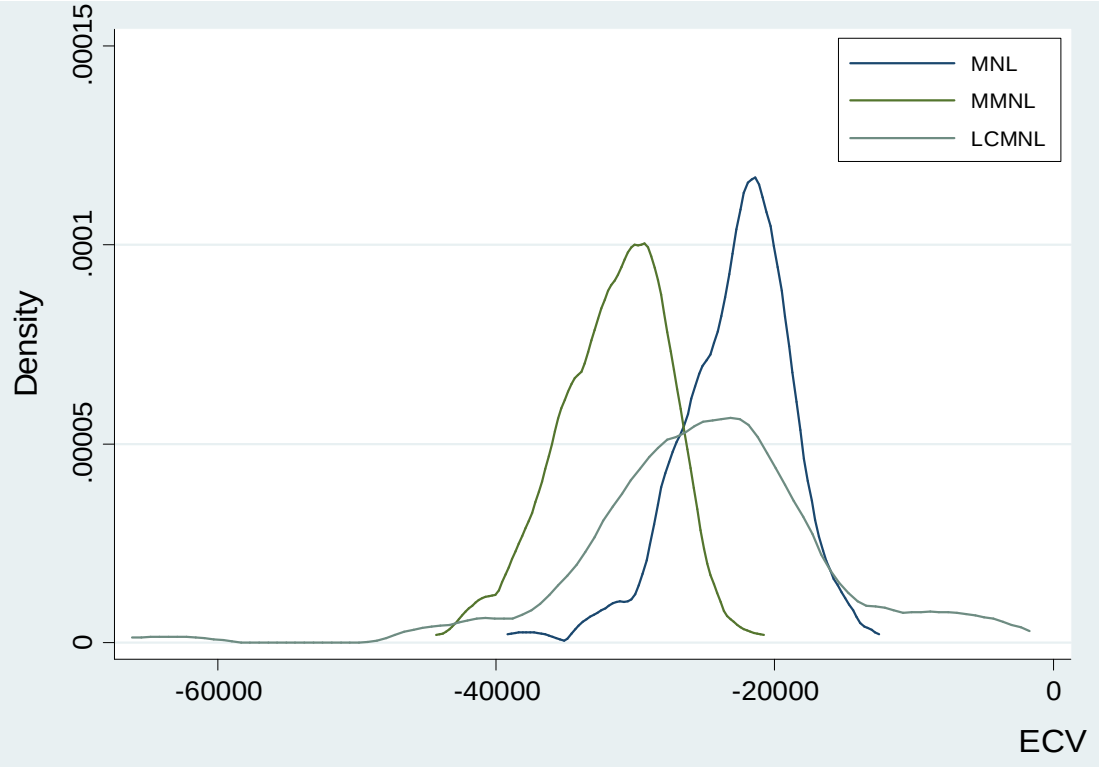
*Note*s: The 95% CIs above are “standard or classical confidence intervals” calculated using 100 bootstrapped estimates of MWTP. This is because accurate, less-erratic and reliable “bootstrap confidence intervals” require replications in the order of 1000, which would have been computationally demanding and time consuming [181]. The confidence intervals reported are therefore not exact.

As with the UK sample, I evaluated the welfare change from reverse engineering a given formulation of drug *C* to a more patient-friendly version *D*. I assumed that both versions of the drug have the same molecule, efficacy and safety profile. Drug *C* is manufactured for intravenous administration in clinical settings, and this mode of drug delivery is associated with "severe" risk of treatment non-compliance and "severe" risk of medication errors. Drug *D* is manufactured for subcutaneous self-administration in non-clinical settings and this mode of drug delivery is associated with "moderate" risk of treatment non-compliance and "moderate" risk of medication errors.

Based on the MNL model, switching from drug *C* to *D* yields an $\widehat{\text{ECV}}$ (per patient per single full treatment course over a year) of -$22,790 (95% CI: -$23,562 to -$22,018). Based on the MMNL model, the $\widehat{\text{ECV}}$ is -$31,601 (95% CI: -$32,374 to -$30,828)[19]. Based on the LCMNL model with two latent classes, and unconditional on class membership, $\widehat{\text{ECV}}$ is -$24,932 (95% CI: -$26,653 to -$23,212).

[19] The price/cost coefficient reported in Table 5.7 for the MMNL model ($\widehat{\beta_p}$) was derived from $\ln(\widehat{\beta_p})$ using the following retransformation: $\widehat{\beta_p} = \exp(\ln(\widehat{\beta_p}) + 0.5\text{SE}^2)$. This, however, assumes the individual $\ln(\widehat{\beta_p})$ recovered from the US dataset are normally-distributed with a constant variance. If this assumption doesn't hold, one obtains a biased and less consistent retransformed price coefficient, as reported in Table 5.7. Therefore, in computing drug *C* → drug *D* welfare change, I used the price coefficient from an unreported MMNL model with a normally-distributed price coefficient ($\widehat{\beta_p}$= 0.000066) as this figure is comparable with the price coefficient from the other models.

Figure 5.4: Sample distribution of ECV estimates



Notes: $\widehat{ECV}$ refers to the welfare change of switching from drug option *C* to *D*. The plots are based on 100 bootstrapped estimates of ECV. This is because estimation of more accurate confidence intervals (that require replications in the order of 1000), would have been computationally demanding and time consuming.

Figure 5.4 shows the kernel density plots of the distribution of $\widehat{ECV}$ derived from the MNL, MMNL and LCMNL models. The overlap of the kernel density plots for the MMNL, LCMNL and MNL models provides some reassurance as to the accuracy of the estimates obtained. The difference between the plots for the LCMNL and MMNL models probably reflects the different assumptions about preference heterogeneity. That the plot for the MNL model overlaps more with that of the LCMNL model confirms that the estimates based on the MNL model are less biased when the IIA assumption holds to an appreciable extent.

I believe that the reason why we see a wider spread of the kernel density plot of $\widehat{ECV}$ derived from the LCMNL model (relative to the MNL or MMNL models) is: because the LCMNL estimates are based on the average coefficients over the two-latent classes (with weights given by the probability of latent-class membership). Hence if the probabilities of class-membership do not remain constant or fixed for each of the 100 bootstrap samples, one

would obtain a lot more variation in $\widehat{ECV}$ than if (fortuitously) the proportion of individuals belonging to each of the latent classes does not vary from one bootstrap sample to the other.

5.5 Discussion

5.5.1 UK Sample

The non-zero MWTP and ECV estimates reported above provide a monetary measure of the "clinical usability" of a drug – where clinical usability has to do with the mode of drug administration as separate from considerations of efficacy, safety and/or value-for-money. Some might consider MWTP and ECV as old-fashioned, redundant metrics of welfare change – arguing that it is better to evaluate predicted choice probabilities for a selected group of products (bundles of attributes). However, such "discrete demand" analyses will not allow us to compare the monetary value of the intangible benefits from making patient-friendly medicines with the monetary savings in drug administration costs. From my analyses, I found that the monetary value of the intangible benefits (from satisfied patients' preferences) is, at least, in the same order of magnitude as savings on the direct monetary costs of resources healthcare providers spend on drug administration.

In addition, the results indicate a strong positive preference for modes of drug administration that are associated with some but not significant risks of adverse events and/or disruptions to patients' daily activities. They also show a positive preference for self-administration of drugs in non-clinical settings – and a negative preference for drug administration in clinical settings or non-clinical settings under the supervision of a qualified healthcare professional. Advances in biopharmaceutical manufacturing such as pre-filled syringes, auto-injectors and pen injectors and other innovations that reduce the risk of adverse events and/or disruptions to daily activities clearly hit with these observations.

This argument, however, assumes that the state of biopharmaceutical manufacturing and formulation science is mature enough to support the desired innovations in making patient-friendly medicines and/or that the profit signals are strong enough to get manufacturers to consider end-user preferences. There might be, of course, practical manufacturing challenges that militate against reverse engineering product option C1 to C2. Nevertheless, the estimates indicate there are substantial benefits from developing patient-friendly drug delivery systems if the underlying formulation and manufacturing science

makes it possible to do so. If these societal benefits are considered important by policy makers, then there is a case for public interventions to encourage manufacturing research in an attempt to achieve the desired goal of producing "clinically usable" medicines and devices.

With the right pricing and reimbursement environment, an additional incentive for manufacturers to consider the preferences of end-users may come from attempts to differentiate products in order to maintain or increase market shares. For any given cohort of patients (consumers), a drug product that closely matches the preferences of the average representative consumer should enjoy higher demand volumes (keeping prices unchanged). Product differentiation along the lines of satisfying end-user preferences for the mode of drug administration may indeed create "brand loyalty" without manufacturers engaging in academic detailing or direct-to-consumer advertising. One would also expect additional demand inducement where manufacturing a drug product in a patient-friendly manner, amplifies the (incremental) health benefits derived from that drug (for example, through reduced risk of adverse events and/or disruptions to patients' daily activities). This is particularly important considering healthcare payers and providers' requirements for estimates of cost-effectiveness from manufacturers to demonstrate product value.

If healthcare payers and providers are willing to pay for the value of the drug delivery mode, and the discounted present value of private producer surplus of developing patient-friendly drug delivery systems (relative to other investment opportunities) is positive, then manufacturers should consider the switch from C1 to C2.

5.5.2 US Sample

For the US sample, my ECV estimates, which provide a monetary value of the welfare gain from making patient-friendly medicines (1) are in the same order of magnitude as the annual acquisition costs ($17,017 to $41,888) for some biologic drugs [182]; and (2) exceed the annual direct monetary costs of administering most drugs (for a single full treatment course). They are likely to fall in the "high end" of the distribution of the direct monetary costs of drug administration reported in Chapter 4, Figure 4.2 on page 88. The results suggest a strong positive preference for modes of drug administration that are associated with some but not significant risks of treatment non-compliance and/ or medication errors. Relative to needle-free routes, there is either a positive preference or

indifference to intravenous and subcutaneous drug delivery; plus a somewhat negative preference for or indifference to intramuscular modes of administration. My results also show a consistent preference for self-administration of drugs in non-clinical settings – and a consistent negative preference for drug administration in clinical settings or non-clinical settings under the supervision of a qualified healthcare professional. For this reason, advances in pharmaceutical manufacturing such as "closed vial access devices" or "closed-system transfer devices" that allow healthcare professionals to safely reconstitute and administer what might be considered hazardous drugs should be encouraged and supported[20].

It might be argued that the sign and magnitude of $\widehat{\text{MWTP}}$ from the LCMNL model for the intravenous and subcutaneous modes of administration is counterintuitive as one would typically expect a positive preference for less-invasive needle-free routes. However, there are sound clinical reasons for a positive $\widehat{\text{MWTP}}$ for intravenous and subcutaneous delivery, relative to needle-free routes (even if the effect is not always statistically significant). Not all drugs, especially in emergency situations, can be given via needle-free routes; and the pharmacokinetic profile of some drugs (for example, those that have a narrow therapeutic index) may be such that intravenous administration is the only or most appropriate route of drug delivery. The positive coefficients of the variables for intravenous and subcutaneous modes of administration are not as surprising as it first seems, considering the coefficient for healthcare professionals who cater for inpatients is the *only* statistically-significant predictor of latent-class membership (all other $\mathbf{Z}$ variables had no statistically significant effects on π_c [equation 20]). In fact, this is consistent with a DCE of French physicians' preferences for intravenous and oral cancer chemotherapy that showed a positive preference for intravenous administration in curative settings as opposed to a positive preference for oral (needle-free) administration in palliative (non-curative) settings[151].

5.5.3 Research limitations

[20] Admittedly, imperfect agency means the choices of healthcare professionals may not necessarily match what patients want or prefer. My DCE results, however, do not require healthcare professionals to act as perfect agents on behalf of patients. The negative coefficients on the attributes "severe risk of medication error" and "severe risk of treatment non-compliance" clearly indicate that patients' health benefits do enter into the utility functions of the sample of healthcare professionals. Patients, especially less-informed passive ones, indirectly obtain the gross welfare benefits measured here via the decisions and actions taken by healthcare professionals.

As with all research, a number of limitations apply to the conclusions and arguments above. First, if preferences change over time, then my reported estimates may no longer be valid albeit one will not expect any dramatic differences from what I have reported in this thesis. A possible avenue for future research is to repeat the analysis here using a panel data of discrete choices collected over time. Second, it might be argued that the reported ECV estimates depend on the price/cost levels chosen. However, Hanley et al.[183] have shown that using different levels for the price attribute may not result in statistically-significant differences in estimates of welfare change albeit there is a possibility that such differences might be significant when the ECV estimates are employed in cost-benefit analyses, for example. See also Slothuus et al.[184]. Considering the near zero price coefficients, I will argue that price/cost levels beyond that chosen for the UK and US DCEs, and any non-linearity in the price-attribute effects, are unlikely to change the arguments above.

Third, some might argue that $\widehat{\mathrm{ECV}}$ derived from the LCMNL model suffers from ecological fallacy – as they are based on the average weighted coefficients over two latent-classes. For that matter (erroneous) conclusions that apply at the aggregate level may not apply at the latent-class level. However, we do not know *a priori* which latent-class a given respondent belongs to – and we cannot assume fixed class membership. Hence the reported $\widehat{\mathrm{ECV}}$, which is unconditional on class membership, is a valid measure of welfare change.

Fourth, the hypothetical choice scenarios presented to US healthcare professionals required that they act as fund-holders taking on payer responsibilities. It might then be argued that the choice data collected may be different from what might have been elicited from actual healthcare payers (who are also agents acting on behalf of patients). On the other hand, payers in their managerial accounting roles rarely administer drugs to patients themselves: they are less familiar with the day-to-day clinical needs of patients. Also, payers are more likely to concern themselves with the direct costs and direct health benefits a drug offers rather than the attributes of drug administration and the associated intangible benefits. On balance, I believe asking healthcare professionals to make discrete drug choices assuming they had financial control of healthcare resources is a better approach compared to a sample population of payers – not to mention the practical problems of identifying such a sample. Note, however, that the US choice data collected was not in a form that allows subgroup analyses of the preferences of doctors and nurses. That the preferences of doctors differ from that of nurses might be worth pursuing in a future DCE study designed to investigate this issue.

Finally, the preference orderings and end-user benefits (ECV) estimated for the US sample of healthcare professionals may not be generalizable to doctors and nurses in the UK because of differences in country-contexts; for example, the role of NICE in determining the mix of drug products in the single-payer UK healthcare system versus the multi-payer environment in the US. Variation in clinical practice styles and other unmeasured or unobserved factors that influence drug choices may also play a role. Nevertheless, the evidence does provide some indication of what attributes healthcare professionals prefer – and the corresponding monetary valuations of the intangible benefits associated with different modes of administering drugs. Recall that time and resource constraints are the reasons why I used a sample of healthcare professionals in the US rather than those in the UK: it was cheaper and quicker for the vendor to conduct the US survey. The more general question of whether preferences for modes of drug administration differ from country to country is an empirical matter that is outside the scope of this thesis. It will require an extensive multicountry DCE to establish whether these differences exist or not.

5.6 Conclusions

In this chapter, I attempted to estimate the monetary value of end-user preferences for a generic set of attributes of different modes of drug administration. For the UK sample of people from the general public, I found a non-trivial marginal willingness-to-pay for drug delivery systems associated with zero or moderate risk of adverse events and/or disruption to patients' daily activities. I also found a high marginal willingness-to-pay for self-administration of drugs in non-clinical settings. In addition, I estimated that the monetary value of making patient-friendly medicines could be as large as the savings on direct monetary costs of drug administration to healthcare payers and providers. For the US sample of healthcare professionals, I found a non-trivial marginal willingness-to-pay for drug delivery systems that are associated with zero or moderate risk of medication errors and/or treatment non-compliance. I also found a high marginal willingness-to-pay for self-administration of drugs in non-clinical settings. In addition, I estimated that the monetary value of making patient-friendly medicines is as large as the annual acquisition costs of some biologic drugs and/or are likely to fall in the "high-end" of the distribution of the direct monetary costs of drug administration.

I will argue that as long as there is recognition of the value of the drug delivery mode (besides the value of drug molecules in improving health outcomes); and the underlying manufacturing science is capable, a patient-centred approach to producing beneficial drugs and drug delivery systems should be encouraged and pursued.

SUMMARY AND RESEARCH RECOMMENDATIONS

6 Putting the Pieces Together

In this thesis, I sought to investigate the relationship between advances in pharmaceutical manufacturing and the costs of delivering healthcare interventions utilizing pharmaceuticals to patients. In the preceding sections of this thesis, I focused on two possible routes by which manufacturing improvements could have an impact on the cost of resources used in delivering healthcare, namely (1) the costs of administering drugs to patients in the event that they fall sick; and (2) costs of acquiring or purchasing drugs that offer positive health benefits net of any concerns about safety and health risks. In line with the objectives listed in Chapter 1, I have been able to: (1) generate, in Chapter 2, empirical evidence on the nature of the relationship between pharmaceutical manufacturing and drug pricing; (2) conduct, in Chapter 3, a systematic review of literature that report estimates of drug administration costs for biologic drugs; (3) develop, in Chapter 4, an algorithm for predicting savings in administration costs from drug reformulation efforts; and (4) conduct, in Chapter 5, a discrete-choice modelling of end-user preferences for difference modes of drug administration.

Put together, the outputs of this thesis indicate there are significant societal benefits from investing in pharmaceutical manufacturing research with the aim of producing new or reformulated drugs that, at least, offer the same stream of health benefits but consume less in terms of healthcare resources. I elucidate on this point in detail below.

6.1 Impact on drug administration costs

Starting with a systematic review of UK literature (that report empirical estimates of the costs of administering biologics), I found that researchers adopted different rules or criteria in deciding what resources costs should be included or excluded in computing estimates of drug administration costs. Consequently, I observed inconsistencies in the magnitude of some of the cost estimates reported in the literature reviewed. To correct these inconsistencies, I first developed a framework to standardize what resource costs should be included in estimates of drug administration costs. This framework makes a distinction between the “physical costs of drug administration”, which refers to the cost of resources spent on physically administering a drug into a patient via one of the body’s orifices; and the

"proximal costs of drug administration", which refers to the resources expended before and after physical administration of a drug to a patient.

Given the observed inconsistencies in what was included or excluded from estimates of drug administration costs, and limitations of my systematic review, I thought it prudent to conduct a *de novo* analysis of the costs of administering a sample of eighteen biologic drugs (none of which are given orally) using the framework mentioned above to careful identify and select what resource costs should be included in my estimates of drug administration costs. From this work, I found that biologics administered subcutaneously or intramuscularly had lower costs of administration than that of biologics administered intravenously. This variation in administration costs appear to be explained by the fact that, whilst most of the resource costs associated with subcutaneous and intramuscular biologics came in the form of proximal costs, intravenous biologics consumed resources in both cost centres. That is, I observed the incidence of both proximal and physical administration costs with intravenous drug delivery. I then explored further the nature of this variation in drug administration costs by running a multivariate ordinary least squares (OLS) regression with a log-transformed dependent variable, i.e., the natural log of administration costs for the sample of biologic drugs chosen. The output from this OLS regression was consistent with alternative regression models that are designed to account for heteroskedasticity, skewness or kurtosis in estimates of drug administration costs or the error residuals from these regressions.

I found that the key parameters driving administration costs were the route of administration (i.e., subcutaneous or intramuscular delivery as opposed to intravenous administration) and the frequency of dosing. From the outputs of my regressions, I was able to develop an algorithm with which one could predict (within reasonable margins of error) the administration costs of biologic drug candidates that are still under development and yet to be launched on the market. The algorithm could be used to estimate savings in the monetary costs of drug administration that could be obtained from reformulating or reverse-engineering an existing drug product or for one that is still under development.

An issue with the research described above is the focus on the monetary costs of drug administration. Whist this is arguably the most important statistic from the perspective of a healthcare payer or provider, one cannot ignore the fact that there are intangible benefits or costs (depending on how one choses to look at it) associated with a given mode of drug administration. For example, a mode of drug administration that incurs the lowest monetary resource costs could carry higher overall societal costs in terms of a mismatch with what

patients (the ultimate end-users) want or prefer. It is useful therefore to estimate in monetary terms what value patients might place on different modes of drug administration using a discrete-choice experiment. Using choice preferences elicited and collected from a sample of people from the general public in the UK, and a sample of healthcare professionals in the US, I estimated a number of discrete-choice regressions that either controlled for unobserved heterogeneity in the choice data collected (assuming homogenous preference structures) or which controlled for heterogeneity in preferences (assuming no heterogeneity in the collective influence of unobserved or unmeasured factors).

Based on the output of the discrete-choice regression (that offered the best fit to the choice data collected), I found the average UK sample respondent preferred modes of drug administration that are associated with some but not a significant risk of adverse events (specific to the mode of administration and separate from those attributable to the drug molecule itself). The average respondent also preferred (1) modes of drug administration that caused less disruptions to patients' daily activities (i.e., those that are more convenient to the patient); and (2) self-administering a drug in non-clinical settings such as the home or appropriate public places when compared with drug administration in clinical settings or self-administration under the supervision of a qualified healthcare professional. I also found that the average US sample respondent preferred self-administration in non-clinical settings as well as modes of drug administration associated with zero or moderate risk of medication errors and/or treatment non-compliance. These preference orderings, however, may not be generalizable to the UK because of variations in clinical practice styles and other institutional or context differences; and the possibility that preferences of actual healthcare payers in the US may be different from that of the healthcare professionals who were asked in the DCE to assume they had financial control over healthcare resources.

Having identified the average preference orderings for the attributes of drug administration, I estimated in monetary terms the gross end-user benefits from reformulating or manufacturing a drug product in a more patient-friendly manner. I found that the magnitude of the gross end-user benefits are in the same order of magnitude as the savings on the direct monetary costs of drug administration, predicted using the algorithm developed in Chapter 4. For the same reasons above, the magnitude of welfare benefits estimated using the US sample of doctors and nurses may not be applicable to the UK setting.

In sum, my research on the impact of manufacturing on drug administration costs suggest that, from either a pure healthcare payer or societal perspective, there are significant

benefits from reformulating drugs in a manner that reduces how much is consumed in terms of healthcare resources and in a manner that matches the preferences of patients (*ex ante* healthy people) expressed directly or indirectly via the decisions and actions taken by healthcare professionals.

6.2 Impact on drug acquisition costs

Under this research stream, I aimed to evaluate empirically any links between pharmaceutical manufacturing innovations and the dynamics in drug markets that have a bearing on affordability of medicines and security of medicine supply. I first developed a theoretical framework that suggests that, depending on the stage of economic lifecycle of drugs, savings from reducing manufacturing costs may be passed on to consumers (via their healthcare payers) in the form of lower prices and/or availability of more drug products to cater for heterogeneous clinical needs.

Using data obtained from the UK NHS over the years 1991-2012, I found evidence to support the notion that, holding all else equal, the number of products (what I referred to as product density) is higher for products that are cheaper and easier-to-make relative to products that are expensive and more difficult-to-make. My findings, however, did not indicate any statistically significant differences in the hazard of exit (i.e., the chances that products are no longer available on the NHS market) for products made with cost-reducing process innovations versus those made with cost-increasing process innovations. The implications of this is: governments could potentially hedge against the risk of medicine shortages by investing in pharmaceutical manufacturing research that seeks to find (increasingly) cheaper ways of making drugs. But this is not all the benefits to be derived from such investments. I also found that variations in drug prices is related to the nature of the underlying manufacturing process. To be precise, the average drop in prices of drugs that are cheaper to make, over time, is roughly twice that of products that are expensive to make – holding all else equal.

The evidence together supports the argument that there are significant societal benefits from improving processes used in drug manufacturing – in terms of lower drug prices, more products available for clinical use, lower drug administration costs and satisfied end-user preferences for the mode of drug administration. Investments in pharmaceutical

manufacturing research should help ensure a more secure supply of affordable, clinically-usable and patient-friendly medicines.

6.3 Contribution

As indicated in Chapter 1, the topic of changes or advances in pharmaceutical manufacturing is often ignored in health-economic literature. I believe therefore that this thesis makes a significant contribution to the existing theoretical and empirical evidence by generating new insights into the societal benefits of improvements in pharmaceutical manufacturing as well as the implications for manufacturability of biologic drug candidates currently in development. The analyses and evidence presented in this thesis, of course, can be updated and expanded on in line with new developments and issues that arise with regards to improving the efficiency of pharmaceutical manufacturing. Granted, this thesis presents a useful starting platform for future work.

This should, however, consider a number of methodological limitations (see below) and estimation issues I identified. First is: the importance of comparing and contrasting the outputs of alternative econometric models derived from the same datasets. This allows one to check the consistency of estimates obtained and identify plausible explanations of deviations from prior expectations. Second, the STATA modules or ado files used for estimating the econometric models selected by a researcher may be problematic with small samples or large datasets. For instance, I found that the STATA modules used for estimating discrete-time cloglog survival models had problems of generating reliable coefficient estimates (on convergence) with a panel data on biologic products spanning the period 1991-2012. On the other hand, the STATA estimator for PGLM/EEE with a power variance function failed to converge with the small sample of 18 biologic drugs I used in developing the administration-cost algorithm. Researchers must take care to identify and consider such (perhaps case-specific) estimation issues when conducting future related work.

6.4 Limitations and future research

A number of limitations apply to the works presented in this thesis. First is: the time period of analysis. This is true especially for the DCE on patients and health-professionals' preferences for different modes of drug administration. Since preferences are not fixed and

may change over time, I will argue that, at least, the two DCEs presented in this thesis are repeated over time. Collecting a panel data on preferences for different modes of drug administration over different time periods may then provide more insights as to the most stable preferred attributes of drug administration.

The second is: the nature of the product sample used in my analyses. For instance, the work on establishing a link between pharmaceutical manufacturing and market conduct used a sample that was skewed by the inclusion of semi-synthetic biologics and the exclusion of existing and emerging cohorts of biologic/ macromolecular therapies. More generalizable evidence can be generated by using more balanced samples of existing and emergent biologic products that are manufactured using pure bioprocessing methods and have reached the later stages of their lifecycles. In the case of developing an administration-cost algorithm, I used a simulated dataset for a sample of 18 biologic drugs, none of which are administered orally. Again, more generalizable evidence can be obtained from expanding the sample (beyond the 18 biologics selected in Chapter 4 to include future oral biologics); and perhaps conducting time-and-motion studies in clinical settings. The latter will yield real-life empirical distributions of the resources costs of administering biologic therapies.

Third is: my thesis did not consider the link between advances in pharmaceutical manufacturing and estimates of cost-effectiveness as measured by NICE and other relevant HTA agencies. The two broad parts of my thesis have clearly demonstrated that advances in drug manufacturing could alter how much of healthcare resources is spent on drug acquisition and administration. This opens up the possibility that advances in drug manufacturing could also have some bearing on cost-effectiveness. The link is as follows. In a typical assessment of the value-for-money offered by drug products, one computes a measure of cost-effectiveness by comparing the discounted present value of lifetime costs associated with a drug with the discounted presented values of health benefits offered by the drug, often measured in terms of quality-adjusted life years. The discounted present value of lifetime costs will be computed as the sum of discounted present values of drug acquisition costs, drug administration costs and the aggregated future healthcare-related costs that are in part determined by the stream of health benefits offered by a drug. Holding all else equal, advances in manufacturing could have a bearing on estimates of cost-effectiveness via the impact on, at least, drug acquisition and drug administration costs.

This means that if the cost-effectiveness of a drug is appraised at different time points along its economic lifecycle, this should bear some relationship to the underlying

manufacturing process employed in making the drug and the alternative treatment comparators to that drug if, over the same time points, changes are made to the underlying manufacturing processes. In fact, one could envisage the possibility of reversals in cost-effectiveness estimates if manufacturing changes contribute to appreciable changes in drug acquisition and drug administration costs.

6.5 Conclusions

Increasing healthcare expenditures have got healthcare payers and providers to consider a number of measures that could help strike a healthy balance between improvements in population health and the amount of resources expended on the provision of good quality medical care. This thesis highlights a less noted tool available to decision makers: investments in pharmaceutical manufacturing research. The value of such investments lies with the links between public health and innovations in pharmaceutical manufacturing – namely lower drug prices, lower resources consumed in administering drugs, ensuring security of supply (to hedge against the risk of medicine shortages) and giving patients, the ultimate end-users, what they want or prefer in terms of the mode of drug administration.

REFERENCES AND APPENDICES

References

1. Kelly C, Mir F. Biological therapies: how can we afford them? BMJ British medical journal 2009; 339(7722):666-669.

2. Grabowski HG, Ridley DB, Schulman KA. Entry and competition in generic biologics. Managerial and Decision Economics 2007; 28(45):439-451.

3. Pauly MV. Drug and vaccine pricing and innovation: what is the story? Managerial and Decision Economics 2007; 28(45):407-413.

4. Vernon JA, Hughen WK, Trujillo AJ. Pharmaceutical manufacturing efficiency, drug prices, and public health: examining the causal links. Drug information journal 2007; 41(2):229-239.

5. Claxton K, Briggs A, Buxton MJ, Culyer AJ, McCabe C, Walker S et al. Value based pricing for NHS drugs: an opportunity not to be missed? BMJ: British Medical Journal 2008; 336(7638):251.

6. Miller DW. Value-Based Pricing. Pharmaceutical Medicine 2012; 26(4):217-222.

7. Mills A, Drummond M. Value for money in the health sector: the contribution of primary health care. Health policy and planning 1987; 2(2):107-128.

8. Drummond MF. Allocating resources. International journal of technology assessment in health care 1990; 6(01):77-92.

9. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health economics 2012; 21(2):145-172.

10. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. Pharmacoeconomics 2008; 26(8):661-677.

11. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ: British Medical Journal 2000; 320(7248):1530.

12. Davidson A, Farid SS. Innovation in Biopharmaceutical Manufacture. BioProcess International 2014; 12:1.

13. Kelly CJ, Mir FA. Economics of biological therapies. BMJ 2009; 339.

14. Cohen J, Stolk E, Niezen M. The increasingly complex fourth hurdle for pharmaceuticals. PharmacoEconomics 2007; 25(9):727-734.

15. Farid SS. Process economic drivers in industrial monoclonal antibody manufacture. In: Gottschalk U, editor. Process Scale Purification of Antibodies. Hoboken, NJ: John Wiley & Sons; 2009. p. 239-261.

16. Moscho A, Hodits RA, Janus F, Leiter JM. Deals that make sense. Nature Biotechnology 2000; 18(7):719-722.

17. Smith DM. Assessing commercial opportunities for autologous and allogeneic cell-based products. Regenerative medicine 2012; 7(5):721-732.

18. Scherer FM. The Link Between Gross Profitability And Pharmaceutical R&D Spending. Health Affairs 2001; 20(5):216-220.

19. Basu P, Joglekar G, Rai S, Suresh P, Vernon J. Analysis of Manufacturing Costs in Pharmaceutical Companies. J Pharm Innov 2008; 3(1):30-40.

20. Lichtenberg F. Sources of U.S. longevity increase, 1960-2001. The Quarterly review of economics and finance 2004; 44(3):369-389.

21. DiMasi JA, Paquette C. The economics of follow-on drug research and development. Pharmacoeconomics 2004; 22(Suppl.2):1-14.

22. Reksulak M, Shughart WF, Tollison RD. Innovation and the opportunity cost of monopoly. Manage Decis Econ 2008; 29(8):619-627.

23. Stigler GJ. The Economies of Scale. Journal of Law and Economics 1958; 1:54-71.

24. Mansfield E, Rapoport J, Romeo A, Wagner S, Beardsley G. Social and Private Rates of Return from Industrial Innovations. The Quarterly Journal of Economics 1977; 91(2):221-240.

25. Rice N, Jones A. Multilevel models and health economics. Health Economics 1997; 6(6):561-575.

26. Cameron AC, Trivedi PK. Microeconometrics Using Stata. Revised Edition ed. Stata Press; 2010.

27. Jones AM, Rice N, d'Uva TB, Balia S. Applied health economics. Routledge; 2013.

28. Jenkins SP. Estimation of discrete time (grouped duration data) proportional hazards models: pgmhaz. Stata Technical Bulletin Reprints , STB 1997; 17(7):109-121.

29. Virabhak S, Sohn W. Drug competition and voluntary exit. Economics Letters 2008; 101(1):34-37.

30. ELLISON SF, SNYDER CM. COUNTERVAILING POWER IN WHOLESALE PHARMACEUTICALS. The Journal of Industrial Economics 2010; 58(1):32-53.

31. Kaufman RL. Decomposing longitudinal from cross-unit effects in panel and pooled cross-sectional designs. Sociological methods & research 1993; 21(4):482-504.

32. Kardos N, Demain A. Penicillin: the medicine with the greatest impact on therapeutic outcomes. Appl Microbiol Biotechnol 2011; 92(4):677-687.

33. Elander RP. Industrial production of beta-lactam antibiotics. Appl Microbiol Biotechnol 2003; 61(5-6):385-392.

34. Buckland B, Josefsberg JO. Biosimilars: Alternative perspectives. Unpublished presentation slides, available on request from the author [2013

35. Hilleman M. Six decades of vaccine development - a personal history. Nature medicine 1998; 4(5s):507-514.

36. Hilleman MR. Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. Vaccine 2000; 18(15):1436-1447.

37. Ulmer JB, Valley U, Rappuoli R. Vaccine manufacturing: challenges and solutions. Nat Biotech 2006; 24(11):1377-1383.

38. Blume S, Geesink I. A brief history of polio vaccines. Science 2000; 288(5471):1593-1594.

39. Blume SS. Lock in, the state and vaccine development: Lessons from the history of the polio vaccines. Research Policy 2005; 34(2):159-173.

40. Tree JA, Richardson C, Fooks AR, Clegg JC, Looby D. Comparison of large-scale mammalian cell culture systems with egg culture for the production of influenza virus A vaccine strains. Vaccine 2001; 19(25):3444-3450.

41. Josefsberg JO, Buckland B. Vaccine process technology. Biotechnol Bioeng 2012; 109(6):1443-1460.

42. US FDA (Food and Drug Administration). FDA approves new seasonal influenza vaccines made using novel technology. Available from URL: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm335891.htm [Accessed: 05/02/2014].

43. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. Vaccine 2007; 25(39):6852-6862.

44. Treanor J. Influenza vaccine - Outmaneuvering antigenic shift and drift. N Engl J Med 2004; 350(3):218-220.

45. Finn A, Bell F. Polio vaccine: is it time for a change? Archives of disease in childhood 1998; 78(6):571-574.

46. Van Hoof J. Manufacturing Issues Related to Combining Different Antigens: An Industry Perspective. Clinical Infectious Diseases 2001; 33(Supplement 4):S346-S350.

47. Fitzpatrick M. MMR Controversy. eLS. John Wiley & Sons, Ltd; 2001.

48. Fitzpatrick M. MMR: risk, choice, chance. British Medical Bulletin 2004; 69(1):143-153.

49. Miller E, Farrington P, Goldracre M, Pugh S, Colville A, Flower A et al. Risk of aseptic meningitis after measles, mumps,and rubella vaccine in UK children. The Lancet 1993; 341(8851):979-982.

50. MHRA (Medicines and Healthcare products Regulatory Agency). Statement from the medicines control agency: Medicines control agency to object to importation of unlicensed single Urabe strain mumps vaccine. Available from URL: http://www.mhra.gov.uk/Safetyinformation/Generalsafetyinformationandadvice/Product-specificinformationandadvice/Vaccinesafety/MMRvaccine/index.htm [Accessed: 15/10/2013]

51. Leland HE. Theory of the firm facing uncertain demand. The American Economic Review 1972; 62(3):278-291.

52. Dana JD. Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure. RAND Journal of Economics 1999; 30(4):632-660.

53. Sikora.K. Paying for cancer care - a new dilemma. Journal of the Royal Society of Medicine 2007; 100(4):166-169.

54. Kelly C, Mir F. Biological therapies: how can we afford them? BMJ 2009; 339(7722):666-669.

55. Aulton ME. Aulton's pharmaceutics: The design and manufacture of medicines. 3rd ed. New York: Churchill Livingstone Elsevier; 2007.

56. Speers M, Bonnano C. Economic aspects of controlled drug delivery. In: Mathiowitz E, editor. Encyclopaedia of controlled drug delivery. 1st ed. New York: Wiley Interscience; 1999. p. 341-347.

57. HCHS. HCHS Pay and Prices Series 2010-2011. Available from URL: http://www.info.doh.gov.uk/doh/finmannsf/af3d43e36a4c8f8500256722005b77f8/276315c0677bf5478025796b00418a4d?OpenDocument [Accessed: 15/11/2012]

58. Drummond M, Aristides M, Davies L, Forbes C. Economic evaluation of standard heparin and enoxaparin for prophylaxis against deep vein thrombosis in elective hip surgery. British Journal of Surgery 1994; 81(12):1742-1746.

59. Le LK, Myon E, Hill S, Riou-Franca L, Scott D, Sidhu M et al. Comparative cost-minimisation of oral and intravenous chemotherapy for first-line treatment of non-small cell lung cancer in the UK NHS system. European Journal of Health Economics 2007; 8(2):145-151.

59. Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. Health Technology Assessment 2006; 10(42):1-248.

60. Jobanputra P, Barton P, Bryan S, Burls A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. Health Technology Assessment 2002; 6(21):1-110.

61. Cummins E, Asseburg C, Punekar YS, Shore E, Morris J, Briggs A et al. Cost-effectiveness of infliximab for the treatment of active and progressive psoriatic arthritis. Value in Health 2011; 14(1):15-23.

62. Woolacott N, Vergel YB, Hawkins N, Kainth A, Khadjesari Z, Misso K et al. Etanercept and infliximab for the treatment of psoriatic arthritis: A systematic review and economic evaluation. Health Technology Assessment 2006; 10(31):September.

63. Bravo VY, Hawkins NS, Claxton K, Asseburg C, Palmer S, Woolacott N et al. The cost-effectiveness of etanercept and infliximab for the treatment of patients with psoriatic arthritis. Rheumatology 2007; 46(11):1729-1735.

64. Dretzke J, Edlin R, Round J, Connock M, Hulme C, Czeczot J et al. A systematic review and economic evaluation of the use of tumour necrosis factor-alpha (TNF-alpha) inhibitors, adalimumab and infliximab, for crohn's disease. Health Technology Assessment 2011; 15(6):2011.

65. Kielhorn A, Porter D, Diamantopoulos A, Lewis G. UK cost-utility analysis of rituximab in patients with rheumatoid arthritis that failed to respond adequately to a biologic disease-modifying antirheumatic drug. Current Medical Research & Opinion 2008; 24(9):2639-2650.

66. Knight C, Hind D, Brewer N, Abbott V. Rituximab (MabThera) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation. [Review] [93 refs]. Health technology assessment (Winchester, England) 2001; 8(37):iii-ixi.

67. Sweetenham J, Hieke K, Kerrigan M, Howard P, Smartt PF, McIntyre AM et al. Cost-minimization analysis of CHOP, fludarabine and rituximab for the treatment of relapsed indolent B-cell non-Hodgkin's lymphoma in the UK. British Journal of Haematology 1999; 106(1):47-54.

68. Cassidy J, Douillard JY, Twelves C, McKendrick JJ, Scheithauer W, Bustova I et al. Pharmacoeconomic analysis of adjuvant oral capecitabine vs intravenous 5-FU/LV in Dukes' C colon cancer: the X-ACT trial. British Journal of Cancer 2006; 94(8):1122-1129.

69. Holmes J, Dunlop D, Hemmett L, Sharplin P, Bose U. A cost-effectiveness analysis of docetaxel in the second-line treatment of non-small cell lung cancer. PharmacoEconomics 2004; 22(9):581-589.

70. Hoyle M, Green C, Thompson-Coon J, Liu Z, Welch K, Moxham T et al. Cost-effectiveness of temsirolimus for first line treatment of advanced renal cell carcinoma. Value in Health 2010; 13(1):61-68.

71. Lewis G, Peake M, Aultman R, Gyldmark M, Morlotti L, Creeden J et al. Cost-effectiveness of erlotinib versus docetaxel for second-line treatment of advanced non-small-cell lung cancer in the United Kingdom. Journal of International Medical Research 2010; 38(1):9-21.

72. Robinson P, von der MH, Bhalla S, Kielhorn A, Aristides M, Brown A et al. Cost-utility analysis of the GC versus MVAC regimens for the treatment of locally advanced or metastatic bladder cancer. Expert Review of Pharmacoeconomics and Outcomes Research 2004; 4(1):27-38.

73. Woolacott N, Bravo VY, Hawkins N, Kainth A, Khadjesari Z, Misso K et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technology Assessment 2006; 10(31):1-258.

74. Bravo VY, Hawkins NS, Claxton K, Asseburg C, Palmer S, Woolacott N et al. The cost-effectiveness of etanercept and infliximab for the treatment of patients with psoriatic arthritis. Rheumatology 2007; 46:1729-1735.

75. Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D et al. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer. Health Technology Assessment 2001; 5(24):1-79.

76. Hoyle M, Rogers G, Moxham T, Liu Z, Stein K. Cost-effectiveness of dasatinib and nilotinib for imatinib-resistant or -intolerant chronic phase chronic myeloid leukemia. Value in Health 2011; 14(8):1057-1067.

77. Lloyd A, Kennedy R, Hutchinson J, Sawyer W. Economic evaluation of highly purified menotropin compared with recombinant follicle-stimulating hormone in assisted reproduction. Fertility and Sterility 2003; 80(5):1108-1113.

78. Martin NK, Vickerman P, Miners A, Foster GR, Hutchinson SJ, Goldberg DJ et al. Cost-effectiveness of hepatitis C virus antiviral treatment for injection drug user populations. Hepatology 2012; 55(1):49-57.

79. Wolowacz SE, Roskell NS, Maciver F, Beard SM, Robinson PA, Plumb JM et al. Economic evaluation of dabigatran etexilate for the prevention of venous thromboembolism after total knee and hip replacement surgery. Clinical Therapeutics 2009; 31(1):194-212.

80. Chilcott J, Lloyd JM, Wight J, Forman K, Wray J, Beverley C et al. A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus-negative. Health Technology Assessment 2003; 7(4):1-72.

81. Fenn P, Gray A, McGuire A. An economic evaluation of universal vaccination against hepatitis B virus. Journal of Infection 1996; 32(3):197-204.

82. Siddiqui MR, Gay N, John Edmunds W, Ramsay M. Economic evaluation of infant and adolescent hepatitis B vaccination in the UK. Vaccine 2011; 29(3):466-475.

83. Jit M, Cromer D, Baguelin M, Stowe J, Andrews N, Miller E. The cost-effectiveness of vaccinating pregnant women against seasonal influenza in England and Wales. Vaccine 2010; 29(1):115-122.

84. McIntosh ED, Conway P, Willingham J, Hollingsworth R, Lloyd A. Pneumococcal pneumonia in the UK: how herd immunity affects the cost-effectiveness of 7-valent pneumococcal conjugate vaccine (PCV). Vaccine 2005; 23(14):1739-1745.

85. Pinot De MA, Edmunds WJ, Breuer J. The cost-effectiveness of antenatal varicella screening with post-partum vaccination of susceptibles. Vaccine 2006; 24(9):1298-1307.

86. Trotter CL, Edmunds WJ. Reassessing the cost-effectiveness of meningococcal serogroup C conjugate (MCC) vaccines using a transmission dynamic model. Medical Decision Making 2006; 26:38-47.

87. Vick S, Cairns J, Urbaniak S, Whitfield C, Raafat A. Cost-effectiveness of antenatal anti-D prophylaxis. Health Economics 1996; 5:319-328.

88. Nuijten MJ, Wittenberg W, Lebmeier M. Cost effectiveness of palivizumab for respiratory syncytial virus prophylaxis in high-risk children: a UK analysis. Pharmacoeconomics 2007; 25(1):55-71.

89. Jit M, Bilcke JB, Mangen M-JJ, Salo H, Melliez H, John Edmunds W et al. The cost-effectiveness of rotavirus vaccination: Comparative analyses for five European countries and transferability in Europe. Vaccine 2009; 27(44):6121-6128.

90. Morris S, Ridley S, Munro V, Christensen MC. Cost effectiveness of recombinant activated factor VII for the control of bleeding in patients with severe blunt trauma injuries in the United Kingdom. Anaesthesia 2007; 62(1):43-52.

91. ISPOR (International Society for Pharmacoeconomics and Outcomes Research). Pharmacoeconomic guidelines around the world. Availabe from URL: http://www.ispor.org/peguidelines/index.asp [Accessed: 12/04/2013]

92. Hughes DA, Tilson L, Drummond M. Estimating Drug Costs in Economic Evaluations in Ireland and the UK. Pharmacoeconomics 2009; 27(8):635-643.

93. Stremersch S, Tellis GJ. Strategic bundling of products and prices: a new synthesis for marketing. Journal of Marketing 2002; 66(1):55-72.

94. Webster CS, Merry AF, Gander PH, Mann NK. A prospective, randomised clinical evaluation of a new safety-orientated injectable drug administration system in comparison with conventional methods. Anaesthesia 2004; 59(1):80-87.

95. Kruse GB, Amonkar MM, Smith G, Skonieczny DC, Stavrakas S. Analysis of costs associated with administration of intravenous single-drug therapies in metastatic breast cancer in a U.S. population. Journal of Managed Care Pharmacy 2008; 14(9):November/December.

96. Harley CR, Frytak JR, Tandon N. Treatment compliance and dosage administration among rheumatoid arthritis patients receiving infliximab, etanercept, or methotrexate. American Journal of Managed Care 2003; 9(6:Suppl):Suppl-43.

97. Dakin HA, Devlin NJ, Odeyemi IAO. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. Health Policy 2006; 77(3):352-367.

98. Cheng MM, Alfonso R, Best JH, Garrison LP, Bruhn D, Veenstra DL. The economic value of reducing medication dosing frequency with drug delivery technologies: an evidence assessment. Journal of Medical Economics 2010; 13(2):193-202.

99. Dretzke J, Edlin R, Round J, Connock M, Hulme C, Czeczot J et al. A systematic review and economic evaluation of the use of tumour necrosis factor-alpha (TNF-alpha) inhibitors, adalimumab and infliximab, for crohn's disease. Health Technology Assessment 2011; 15(6).

100. Iveson TJ, Hickish T, Schmitt C, Van CE. Irinotecan in second-line treatment of metastatic colorectal cancer: improved survival and cost-effect compared with infusional 5-FU. European Journal of Cancer 1999; 35(13):1796-1804.

101. Loveman E, Jones J, Hartwell D, Bird A, Harris P, Welch K et al. The clinical effectiveness and cost-effectiveness of topotecan for small cell lung cancer: a systematic review and economic evaluation. Health Technology Assessment 2010; 14(19):1-178.

102. Starling N, Tilden D, White J, Cunningham D. Cost-effectiveness analysis of cetuximab/irinotecan vs active/best supportive care for the treatment of metastatic colorectal cancer patients who have failed previous chemotherapy treatment. British Journal of Cancer 2007; 96(2):206-212.

103. Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N. Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes. Health Technology Assessment 2004; 8(43).

104. Cummins E, Royle P, Snaith A, Greene A, Robertson L, McIntyre L et al. Clinical effectiveness and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes: systematic review and economic evaluation. Health Technology Assessment 2010; 14(11).

105. Cummins E, Asseburg C, Prasad M, Buchanan J, Punekar YS. Cost effectiveness of golimumab for the treatment of active psoriatic arthritis. European Journal of Health Economics 2012; 13(6):801-809.

106. Palmer AJ, Roze S, Valentine WJ, Smith I, Wittrup-Jensen KU. Cost-effectiveness of detemir-based basal/bolus therapy versus NPH-based basal/bolus therapy for Type 1 diabetes in a UK setting: an economic analysis based on meta-analysis results of four clinical trials. Current Medical Research and Opinion 2004; 20(11):1729-1746.

107. Rogers G, Hoyle M, Thompson CJ, Moxham T, Liu Z, Pitt M et al. Dasatinib and nilotinib for imatinib-resistant or -intolerant chronic myeloid leukaemia: a systematic review and economic evaluation. Health Technology Assessment 2012; 16(22):1-410.

108. Mangtani P, Hall AJ, Normand CE. Hepatitis B vaccination: the cost effectiveness of alternative strategies in England and Wales. Journal of Epidemiology and Commmunity Health 1995; 49(3):238-244.

109. Lowin J, Bergman A, Chaudhuri KR, Findley LJ, Roeder C, Schifflers M et al. A cost-effectiveness analysis of levodopa/carbidopa intestinal gel compared to standard care in late stage Parkinson's disease in the UK. Journal of Medical Economics 2011; 14(5):584-593.

110. Walleser S, Ray J, Bischoff H, Vergnenegre A, Rosery H, Chouaid C et al. Maintenance erlotinib in advanced nonsmall cell lung cancer: cost-effectiveness in EGFR wild-type across Europe. ClinicoEconomics and Outcomes Research 2012; 4(1):269-275.

111. Danzon PM, Wang YR, Wang L. The impact of price regulation on the launch delay of new drug - evidence from twenty-five major markets in the 1990s. Health Economics 2005; 14(3):269-292.

112. Tetteh E, Morris S. Systematic Review of Drug Administration Costs and Implications for Biopharmaceutical Manufacturing. Appl Health Econ Health Policy 2013;1-12.

113. Dranove D. Measuring costs. In: Sloan FA, editor. Valuing health care: Costs, benefits, and effectiveness of pharmaceuticals and other medical technologies. Cambridge University Press; 1995. p. 61-75.

114. Arrow KJ. Uncertainty and the Welfare Economics of Medical Care. The American Economic Review 1963; 53(5):941-973.

115. Dardanoni V, Wagstaff A. Uncertainty and the demand for medical care. Journal of health economics 1990; 9(1):23.

116. Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. Statist Med 2004; 23(8):1311-1331.

117. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford university press; 2006.

118. Duan N. Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association 1983; 78(383):605-610.

119. Manning W. The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of health economics 1998; 17(3):283-295.

120. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? Journal of health economics 2001; 20(4):461-494.

121. Wooldridge JM. Introductory econometrics: a modern approach. Cengage Learning; 2012.

122. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. Journal of health services research & policy 2004; 9(4):197-204.

123. Moran JL, Solomon PJ, Peisach AR, Martin J. New models for old questions: generalized linear models for cost prediction. Journal of evaluation in clinical practice 2007; 13(3):381-389.

124. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. Journal of health economics 2005; 24(3):465-488.

125. Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. Biostatistics 2005; 6(1):93-109.

126. Basu A. Extended generalized linear models: simultaneous estimation of flexible link and variance functions. Stata Journal 2005; 5(4):501.

127. Kennedy PE. Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations. The American Economic Review 1981; 71(4):801.

128. Jan van Garderen K, Shah C. Exact interpretation of dummy variables in semilogarithmic equations. Econometrics Journal 2002; 5(1):149-159.

129. Bittner B, Richter W, Hourcade-Potelleret F, McIntyre C, Herting F, Zepeda M et al. Development of a Subcutaneous Formulation for Trastuzumab - Nonclinical and Clinical Bridging Approach to the Approved Intravenous Dosing Regimen. Arzneimittelforschung 2012; 62(09):401-409.

130. Wynne C, Harvey V, Schwabe C, Waaka D, McIntyre C, Bittner B. Comparison of Subcutaneous and Intravenous Administration of Trastuzumab: A Phase I/Ib Trial in Healthy Male Volunteers and Patients With HER2-Positive Breast Cancer. J Clin Pharmacol 2013; 53(2):192-201.

131. Samanta K, Moore L, Jones G, Evason J, Owen G. PCN39 Potential Time and Cost Savings with Herceptin (Trastuzumab) Subcutaneous (SC) Injection Versus Herceptin Intravenous (IV) Infusion: Results from Three Different English Patient Settings. Value in Health 2012; 15(7):A415.

132. De Cock E, Tao S, Urspruch A, Knoop A. RU2 Potential Time Savings With Trastuzumab Subcutaneous (SC) Injection Versus Trastuzumab Intravenous (IV) Infusion: Results from Interviews Conducted As Part of a Time-and-Motion Study (T&M) Across 17 Sites. Value in Health 2012; 15(7):A280.

133. de La Horrie GF. Making Biologic Drugs More Affordable. Drug Discovery and Development Magazine. 8-10-2013.

134. Eisenstein M. Something new under the skin. Nature biotechnology 2011; 29(2):107-109.

135. Parker SE. Pharmacoeconomics of Intravenous Drug Administration. Pharmacoeconomics 1992; 1(2):103-115.

136. Dychter SS. Subcutaneous Drug Delivery: A Route to Increased Safety, Patient Satisfaction, and Reduced Costs. Journal of infusion nursing 2012; 35(3):154.

137. Chess R. Economics of drug delivery. Pharmaceutical Research 1998; 15(2):172-174.

138. Knowledge Transfer Network. The future of high value manufacturing in the UK: Pharmaceutical, biopharmaceutical & medical device sectors. Available from URL: https://connect.innovateuk.org/documents/3112383/10498355/Health+KTN+-+The+future+of+High+Value+Manufacturing+in+the+UK+Pharmaceutical%20Biopharmaceutical+&+Medical+Device+Sectors+-+2013/738f33cc-a53f-417c-910f-339f4727489f [Accessed: 29/10/2014]

139. Tetteh E, Morris S. Evaluating the administration costs of biologic drugs: development of a cost algorithm. Health Economics Review 2014; 4(1):26.

140. Spence M. Product differentiation and welfare. The American Economic Review 1976;407-414.

141. Anderson SP, De Palma A. The logit as a model of product differentiation. Oxford Economic Papers 1992;51-67.

142. Besanko D, Perry MK, Spady RH. The logit model of monopolistic competition: Brand diversity. The Journal of Industrial Economics 1990;397-415.

143. Besanko D, Gupta S, Jain D. Logit demand estimation under competitive pricing behavior: An equilibrium framework. Management Science 1998; 44(11-part-1):1533-1547.

144. Dubë C, Chintagunta P, Petrin A, Bronnenberg B, Goettler R, Seetharaman PB et al. Structural Applications of the Discrete Choice Model. Marketing letters 2002; 13(3):207-220.

145. Giacomo MD. EMPIRICAL ANALYSIS OF MARKETS WITH DIFFERENTIATED PRODUCTS: THE CHARACTERISTICS APPROACH. Giornale degli Economisti e Annali di Economia 2004; 63 (Anno 117)(2):243-288.

146. Lancaster KJ. A New Approach to Consumer Theory. Journal of Political Economy 1966; 74(2):132-157.

147. Lancaster K. Socially optimal product differentiation. The American Economic Review 1975;567-585.

148. Baumol WJ. Calculation of optimal product and retailer characteristics: The abstract product approach. The Journal of Political Economy 1967;674-685.

149. Manski CF. The structure of random utility models. Theory and decision 1977; 8(3):229-254.

150. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P., editor. Frontiers in Econometrics. New York: Academic Press; 1974. p. 105-142.

151. Benjamin L. Physicians' preferences for prescribing oral and intravenous anticancer drugs: A Discrete Choice Experiment. European journal of cancer 2012; 48(6):912.

152. Augustovski F, Beratarrechea A, Irazola V, Rubinstein F, Tesolin P, Gonzalez J et al. Patient preferences for biologic agents in rheumatoid arthritis: a discrete-choice experiment. Value in Health 2013; 16(2):385-393.

153. Huynh TK, Oÿstergaard A, Egsmose C, Madsen OR. Preferences of patients and health professionals for route and frequency of administration of biologic agents in the treatment of rheumatoid arthritis. Patient preference and adherence 2014; 8:93.

154. Parker SE. Pharmacoeconomics of intravenous drug administration. PharmacoEconomics 1992; 1(2):103-115.

155. Dychter SS. Subcutaneous drug delivery: a route to increased safety, patient satisfaction, and reduced costs. Infus Nurs 2012; 35(3):154.

156. McDowell SE, Mt-Isa S, Ashby D, Ferner RE. Where errors occur in the preparation and administration of intravenous medicines: a systematic review and Bayesian analysis. Quality and Safety in Health Care 2010; 19(4):341-345.

157. Keers RN, Williams SD, Cooke J, Ashcroft DM. Prevalence and nature of medication administration errors in health care settings: a systematic review of direct observational evidence. Annals of Pharmacotherapy 2013; 47(2):237-256.

158. Farroni JS, Zwelling L, Cortes J, Kantarjian H. Saving Medicare through Patient-Centered Changes - The Case of Injectables. N Engl J Med 2013; 368(17):1572-1573.

159. Zwerina K, Huber J, Kuhfeld WF. A general method for constructing efficient choice designs. Durham, NC: Fuqua School of Business, Duke University 1996.

160. Kuhfeld WF, Tobias RD, Garratt M. Efficient Experimental Design with Marketing Research Applications. Journal of Marketing Research 1994; 31(4):545-557.

161. Huber J, Zwerina K. The Importance of Utility Balance in Efficient Choice Designs. Journal of Marketing Research 1996; 33(3):307-317.

162. Kuhfeld WF, Tobias RD. Large factorial designs for product engineering and marketing research applications. Technometrics 2005; 47(2).

163. Kuhfeld WF. Marketing research methods in SAS experimental design, choice, conjoint, and graphical techniques. 2009.

164. McFadden DL. Econometric analysis of qualitative response models. 1984.

165. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. Pharmacoeconomics 2008; 26(8):661-677.

166. Orme B. Sample size issues for conjoint analysis studies. Sawthooth Software Research paper Series Squim , WA , USA: Sawthooth Software Inc 1998.

167. Rose JM, Bliemer MC. Sample size requirements for stated choice experiments. Transportation 2013; 40(5):1021-1041.

168. Bridges JFP. Conjoint analysis applications in health - a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health 2011; 14(4):403.

169. LOUVIERE JJ, Islam T, Wasi N, Street D, Burgess L. Designing discrete choice experiments: Do optimal designs come at a price? Journal of Consumer Research 2008; 35(2):360-375.

170. Swait J. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational behavior and human decision processes 2001; 86(2):141.

171. de Bekker-Grob EW, Rose JM, Bliemer MC. A Closer Look at Decision and Analyst Error by Including Nonlinearities in Discrete Choice Models: Implications on Willingness-to-Pay Estimates Derived from Discrete Choice Data in Healthcare. PharmacoEconomics 2013; 31(12):1169-1183.

172. Hole AR. Estimating mixed logit models using maximum simulated likelihood. Stata J 2007; 7(3):388-401.

173. Pacifico D. Fitting nonparametric mixed logit models via expectation-maximization algorithm. Stata Journal 2012; 12(2).

174. Bech M, Gyrd-Hansen D. Effects coding in discrete choice experiments. Health Econ 2005; 14(10):1079-1083.

175. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ 2007; 16(8):827-840.

176. Lancsar E, Savage E. Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Econ 2004; 13(9):901-907.

177. Ryan M. Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1). Health Econ 2004; 13(9):909-912.

178. Santos Silva JMC. Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (2). Health Econ 2004; 13(9):913-918.

179. Trajtenberg M. The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners. Journal of Political Economy 1989; 97(2):444-479.

180. Hausman J, McFadden D. Specification tests for the multinomial logit model. Econometrica : Journal of the Econometric Society 1984;1219-1240.

181. Efron B. Better bootstrap confidence intervals. Journal of the American Statistical Association 1987; 82(397):171-185.

182. Wu N, Shah N, Harrison DJ. Cost of biologics per treated patient across immune-mediated inflammatory disease indications in a pharmacy benefit management setting: a retrospective cohort study. Clinical therapeutics 2014; 36(8):1231-1241.

183. Hanley N, Adamowicz W, Wright RE. Price vector effects in choice experiments: an empirical test. Resource and energy economics 2005; 27(3):227-234.

184. Slothuus Skjoldborg U, Gyrd-Hansen D. Conjoint analysis. The cost variable: an Achilles' heel? Health Econ 2003; 12(6):479-491.

Appendices

Appendix 2A: Some useful background information

The influenza vaccines highlight the fact that, for efficacy or safety reasons, one might observe a shift from cost-increasing to cost-reducing process innovation (and vice versa). The traditional production for making influenza vaccines relies on an adequate supply of chicken eggs (apparently one to two eggs is needed to make *one dose* of the vaccine). Although this egg-based process has benefited from decades of industrial experience and a good safety profile; it lacks flexibility for manufacturing scale-up. The process is labour intensive, requires complex automation with relatively long production times – approximately 6-9 months. This has generated a growing interest in a cell-based culture processes for manufacturing influenza vaccines as this shortens production time by about 10 weeks [35;36;40]. It is worth mentioning that a new influenza vaccine Flublok®, currently only licensed for use in the US, is made in baculovirus-insect cell system using recombinant DNA technology to bypass the reliance on adequate supply of chicken eggs or availability of the influenza virus [41],[42]. This leverage from speedy scale-up of manufacturing, in (mammalian) cell culture, of vaccines offering broad protection against heterovariant virus strains is considered an important solution to the mismatch of influenza vaccine efficacy with circulating viruses due to antigenic drift (the development of variants of a given virus strain) and antigenic shift (emergence of new viruses that had never been in human circulation or last affected humans decades ago) [43;44].

Poliomyelitis (polio) vaccines provide an example of how market structure and conduct can be altered by the interplay between manufacturing process innovations and shifting public policy and regulations. Notably, the 1990's saw an increase in public-health calls for a switch from OPV to an improved or enhanced IPV for safety reasons related to the rare but present risk of vaccine-associated paralytic polio (VAPP) and vaccine-derived polioviruses (VDPV). These calls had to contend with the fact that OPV provides external community benefits whilst IPV only offers private protection for the immunized individual. The exclusive concentration and shift of demand from OPV to (enhanced) IPV in the UK in 2004 can thus be seen as a trade-off between the associated risk of VAPP or VDPV and the marginal external community benefits as the incidence and/or prevalence of polio in the UK fell close to zero. Indeed, of 28 confirmed cases of polio in the UK between 1985 and 1994,

19 were VAPP, six were acquired abroad and the rest were of unknown origin [45]. In the case of a disease epidemic, however, OPV is the preferred option as the marginal external community benefits then outweigh the rare risks of VAPP or VDPV.

Measles-Mumps-Rubella (MMR) vaccines provide an example of how manufacturing challenges interplay with random historical events to affect market structure and conduct. For these vaccines, the usual risks and complexities of manufacturing combination vaccines apply: any change to the manufacturing process of any one of the component vaccines (for example, the move towards preservative-free vaccines to reduce the risk of autism and other neurodevelopment disorders) would affect how the whole combination product is manufactured [46]. But beyond these risks, the period after the MMR vaccines were introduced into the UK market (in 1988) can be characterised by refinements of the production process to reduce COG and demand uncertainty created by the outbreak of the MMR controversy, in 1998, about the possible risk of autism with MMR vaccines – the so-called "Wakefield Hoax". This led to a dramatic fall in uptake of MMR vaccines, which reached a low in 2004. MMR vaccine uptake only began to rise after this period, and continued after the controversy reached its conclusion in 2010 [47;48]. I consider the demand uncertainty created by the MMR controversy as a random 'historical' event that did not average out. But before the Wakefield Hoax, there were *proven* safety concerns that MMR vaccines containing the Urabe strain of the mumps virus was associated with a risk of aseptic meningitis. This led to the withdrawal of Urabe-containing MMR vaccines from use in the UK in 1992. The MMR vaccines currently in use are those containing the Jeryl-Lynn strain of the mumps virus, which after many years of worldwide experience is not thought to be associated with aseptic meningitis [49; 50].

Appendix 2B: Graphical analyses

Figure B1: Product prices, density and number of firms – Penicillins

Figure B2: Product prices, density and number of firms – Hepatitis B vaccines

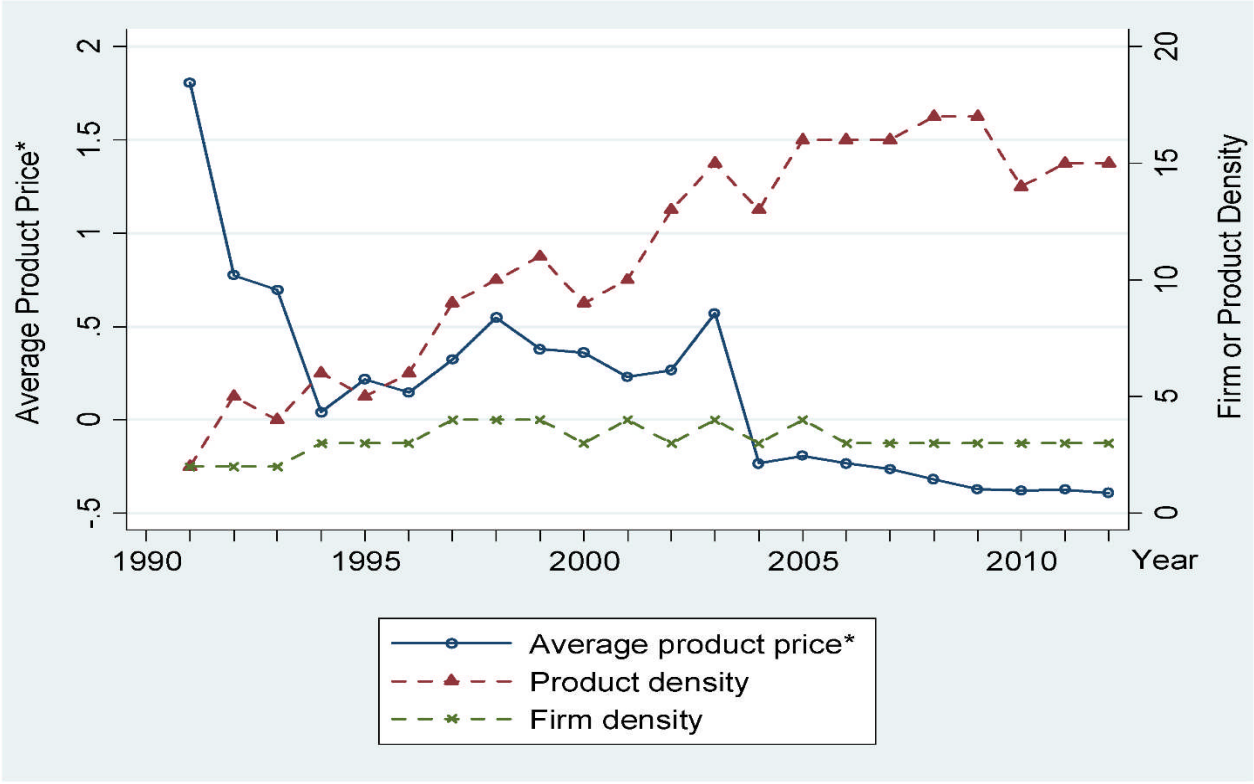


Figure B3: Product prices, density and number of firms – Influenza vaccines

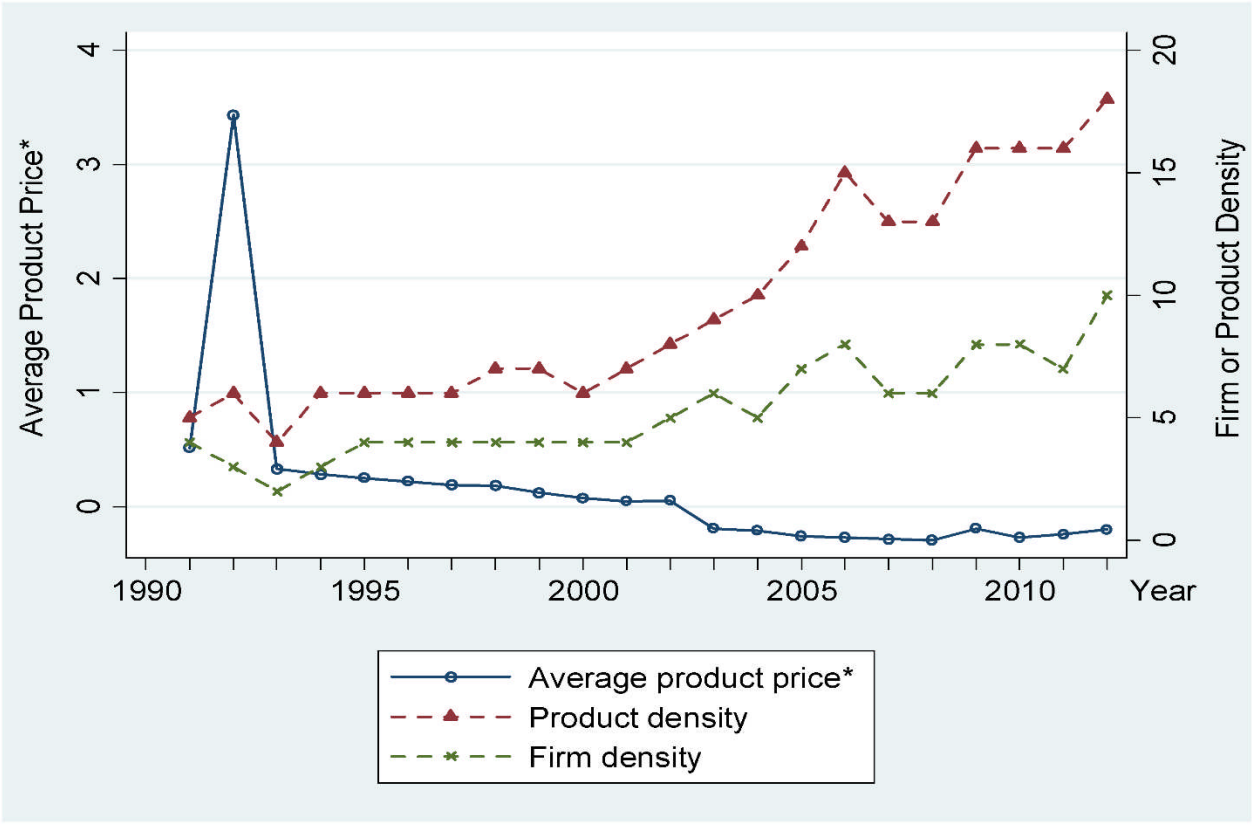


Figure B4: Product prices, density and number of firms – Polio vaccines

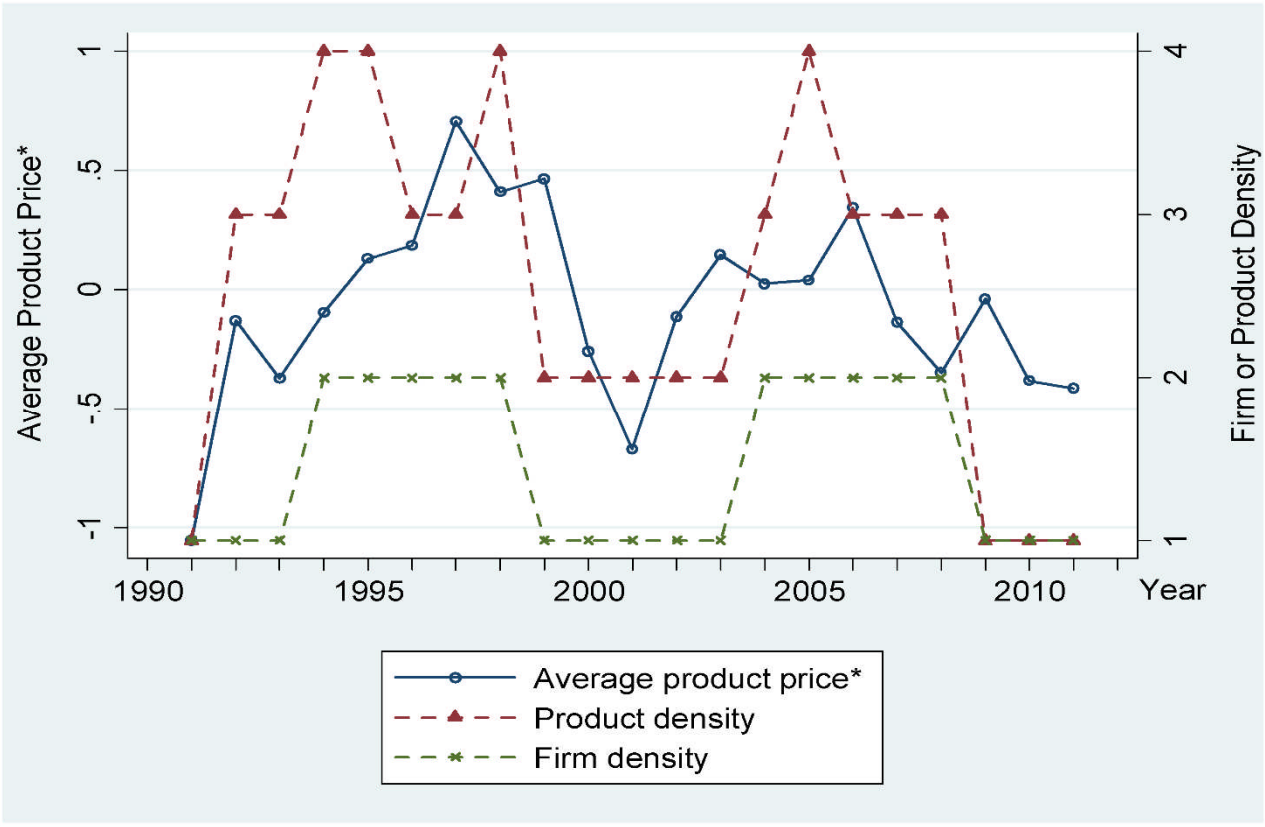
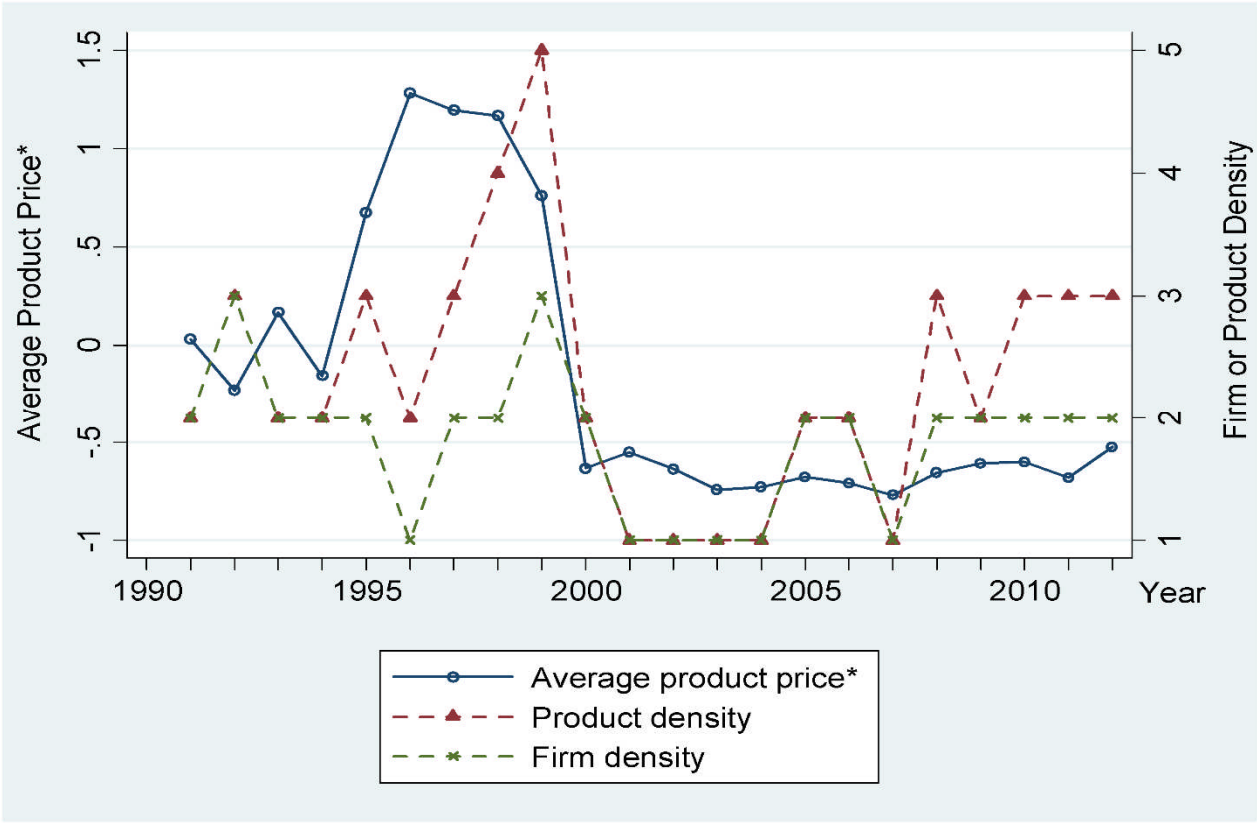


Figure B5: Product prices, density and number of firms – MMR vaccines



Descriptive

Figures B1 to B5 above depict for product class (therapeutic market) over the years 1991-2012: (1) the number of products, what I call product density; (2) firm density, i.e., the number of manufacturers whose identity are recorded in the HSCIC dataset; and (3) the average of *standardized* product prices, indicated by a star (*).

The penicillins (Figure B1) exhibit a skewed inverted U-shaped distribution in product density over time with a similar trend observed for firm density. The initial time period is characterised by entry of products and firms which is then followed by a shakeout period where there is a steady decline in the number of products and firms. Average product price for the penicillins also follows a similar trend albeit there are some erratic (up and down) movements in average product price – due to: (1) the absence of data in the year 1991 followed by data availability in 1992 for some of the more expensive antipseudomonal penicillins; and (2) changes in product mix from exits of old products and introduction of new products that carry a price premium. The general downward trend in average product prices,

firm density and product density, nevertheless, suggests that competitive pressures are fairly functional and effective within the penicillin market.

For the hepatitis B and influenza vaccines (Figures B2 and B3 respectively), we do not observe an inverted U-shaped distribution for product density or firm density over time. In the case of hepatitis B vaccines, I observed a virtually flat firm-density curve whilst the number of products (single-component hepatitis B vaccines or in combination with hepatitis A) rises steadily. One observes a downward trend in prices albeit there are price hikes over some time periods due to the introduction of different formulations and dose strengths of hepatitis-B-containing vaccines that carry a price premium. In the case of influenza vaccines, firm density follows the upward trend in product density, which at the same time is associated with a downward trend in average product prices. The sudden hike and dip in prices from 1991 to 1993 is, however, not due to vaccine supply shortage but changes in product mix (the introduction of an expensive version of Influvac® that exited the market in 1993). I consider that the downward trend in average product price for the hepatitis B and influenza vaccines is not just reflective of functional competitive pressures at play but also consistent with predictions that overcrowded therapeutic markets with a higher number of products and firms under declining prices cannot be sustained without efforts to lower COG.

For the polio and MMR vaccines (Figures B4 and B5), one observes more erratic trends in average product prices, product density and firm density. In the case of polio vaccines, I observed the emergence of a "single-product single-supplier" equilibrium, which can be attributed to public health calls for a switch from OPV to IPV and subsequently the exclusive concentration of NHS demand on IPV. The trend in product and firm density observed for the MMR vaccines is consistent with the withdrawal of (or more accurately the recommendation not to purchase) vaccines containing the Urabe strain of the mumps virus in 1992 and the emergence/outbreak of the Wakefield Hoax in 1998 and its resolution in 2010. Note, in particular, that the drastic dip in firm and product density around year 2000, which persists until after the year 2004. Likewise average product price (under the uncertain demand created) falls steadily over that time period. It is only after 2010 that a 'stable equilibrium' of three products, two suppliers emerges. Note that the sharp hike in average product price for the MMR vaccines between 1991 and 1992 is due to a change in product mix, specifically the introduction of an imported product carrying a price premium. Also, while the dip in average product price and fall in product or firm density may appear inconsistent with the opportunity

created for oligopolistic pricing, it is consistent with firm behaviour under uncertain demand [51;52].

I speculate that a 'riskless' demand curve for MMR vaccines existed before the Wakefield Hoax and this demand curve was largely price-inelastic at prices above marginal/ incremental production costs. The MMR controversy expanded the price-elastic segments in what became a 'risky' demand curve, leading to a fall in the profit-maximizing price or more accurately a fall in the price-cost margin. Indeed, given uncertain demand and fixed costly investments in capacity, a rational manufacturer will offer 'low' short-run prices that are, at least, higher than the average variable costs of production or supply – with the expectation that products will sell with a higher probability. The fact that prices begun to rise after year 2010 (when the controversy was resolved) lends some support to this conjecture.

Caveats

The "prices" used were derived from drug acquisition costs measured by the net-ingredient cost of prescription items dispensed within the NHS. A prescription item refers to a single provision of a medicine on a prescription form or script, i.e., if a prescription form includes three medicines, the number of items is three. Using the net-ingredient-cost (NIC) per item as a price proxy is thus confounded by heterogeneity in pack sizes, dosage forms, dose strengths and duration of treatment.

I believe, this heterogeneity bias will cancel out if the two imprecise measures of price and volume are used together but in isolation, NIC-per-item will inflate product prices whilst the number of items on prescription scripts will underestimate demand volumes. A possible more-homogenous unit of analysis is defined daily doses (DDDs) but then there are no DDDs for vaccines. Therefore, in Figures B1-B5, product prices are based on the dimensionless average of standardized NIC-per-item to minimize any heterogeneity bias.

Note also that with the exception of polio vaccines, firm density is computed as the number of manufacturers or suppliers whose identity is known or recorded in the HSCIC dataset. It excludes manufacturers or suppliers of products described as "imported". Using only the number of *recorded* manufacturers, however, meant that the firm-density curve for polio vaccines in some time periods centred on zero. Notably, the shift towards a "single-product single-supplier" equilibrium would have been associated with zero number of firms in Figure

B4. Since somebody will have to supply the vaccines, I corrected this by adding one to the number of recorded firms supplying polio vaccines in each year.

Appendix 2C: Regression analysis on product density

Dependent variable: ln(N_PRODUCTS)					
Model specified:	Fixed-effects model	Random-effects model	Population-averaged model	Mixed-effects linear model	
Coefficients:	$\hat{\beta}_{FE}$(SE)	$\hat{\beta}_{RE}$(SE)	$\hat{\beta}_{PA}$(SE)	$\hat{\beta}_{ME}$(SE)[a,c]	$\hat{\beta}_{ME}$(SE)[b,c]
COSTREDINNOV	1.1672 (0.421)*	0.9324 (0.258)***	1.1418 (0.396)**	1.1522 (0.059)***	0.2952 (0.086)**
YRSINDEXP	0.0093 (0.003)*	0.0211 (0.005)***	0.0105 (0.003)**	0.0100 (0.001)***	0.0008 (0.001)
COSTREDINNOV*YRSINDEXP	-0.0303 (0.008)**	-0.0261 (0.007)***	-0.0298 (0.008)***	-0.0300 (0.001)***	-0.0049 (0.002)**
PRODUCTAGE	-0.0008 (0.001)	-0.0011 (0.001)	-0.0008 (0.001)	-0.0008 (0.001)	0.0002 (0.001)
PRODUCTAGE^2	0.00001 (0.0004)	0.00004 (0.0004)	0.00001 (0.0004)	0.00001 (0.0004)	-0.00002 (0.0004)
ln(N_FIRMS)	0.7514 (0.058)***	0.7722 (0.050)***	0.7539 (0.056)***	0.7529 (0.013)***	0.6593 (0.014)***
IMPORTS	0.0523 (0.018)**	0.0529 (0.018)**	0.0524 (0.018)**	0.0524 (0.010)***	0.0552 (0.0093)***
ln(DEMANDVOL)	0.0001 (0.0001)	-0.0002 (0.0002)	0.0001 (0.0001)	0.0001 (0.0003)	0.0002 (0.0003)
ln(PRODUCTPRICE)	-0.0003 (0.0005)	-0.0006 (0.0004)	-0.0003 (0.0004)	-0.0003 (0.001)	0.0001 (0.001)
A	2.0655 (0.240)***	1.7805 (0.222)***	2.0350 (0.234)***	2.0474 (0.063)***	2.0735 (0.060)***
B	1.2527 (0.106)***	1.4297 (0.106)***	1.2702 (0.103)***	1.2631 (0.021)***	1.1549 (0.020)***
C	0.5422 (0.089)***	0.5653 (0.087)***	0.5440 (0.088)***	0.5433 (0.019)***	0.6123 (0.018)***
D	0.7832 (0.101)***	0.7664 (0.093)***	0.7818 (0.100)***	0.7824 (0.021)***	0.6693 (0.020)***
Intercept: $\widehat{\beta_0}$(SE)	0.1238 (0.122)	-0.2177 (0.144)	0.1169 (0.117)	0.1323 (0.049)**	0.4635 (0.066)***
$(\widehat{\sigma_{\mu 0}}, \widehat{\sigma_{\epsilon}}; \widehat{\sigma_{\mu 1}})$	(0.177, 0.078)	(0.027, 0.078)	–	(0.167, 0.078)	(0.265, 0.070; 0.252)
Covar$(\widehat{\mu_{0t}}, \widehat{\mu_{1t}})$, corr$(\hat{\beta}_0, \hat{\beta}_{1t})$	–	–	–	–	-0.0648, -0.9701
Variance partition coefficient	0.8362	0.1052	–	0.8213	0.4596
Overall R-squared	0.9769	0.9884	–	–	–
Corr$(\mathbf{X}_{it}, \widehat{\mu_{0t}})$	-0.4833	–	–	–	–
Log-likelihood (AIC)	4221.6 (-8417.1)	–	–	4137.3 (-8242.6)	4510.4 (-8984.8)

Notes: SE = heteroskedastic-robust standard errors; ***p < 0.001; **p < 0.01; *p < 0.05; !p < 0.10; [a] Without a random-slope for COSTREDINNOV; [b] With a random-slope for COSTREDINNOV; [c] Log-likelihood ratio (LR) test for a comparison with the equivalent single-level model was statistically significant. LR test for a comparison of (a) and (b) indicates statistically-significant evidence that the effect of COSTREDINNOV is not fixed: it varies across different time periods; AIC = Akaike Information Criterion.

Appendix 2D: Regression analysis on hazard of product exit – without unobserved heterogeneity

Dependent variable: Product exit								
Model specified:	Cox-PH model		Weibull model		Gompertz model		Cloglog PH model	
Coefficients:	$\hat{\beta}_1$(SE)	HR	$\hat{\beta}_1$(SE)	HR	$\hat{\beta}_1$(SE)	HR	$\hat{\beta}_1$(SE)	HR
COSTREDINNOV	-3.8372 (3.693)	0.0216	-3.4433 (4.001)	0.0320	-3.0878 (3.930)	0.0456	0.4095 (3.302)	1.5061
YRSINDEXP	0.0732 (.059)	1.0759	0.0828 (0.064)	1.0863	0.0801 (0.063)	1.0834	0.0120 (0.037)	1.0121
COSTREDINNOV*YRSINDEXP	0.0629 (0.084)	1.0650	0.0532 (0.091)	1.0547	0.0503 (0.090)	1.0515	0.0454 (0.070)	1.0465
PRODUCTAGE	-0.0745 (0.018)***	0.9282	-0.0916 (0.019)***	0.9124	-0.0861 (0.019)***	0.9175	0.2363 (0.030)***	1.2665
ln(N_FIRMS)	2.9547 (0.624)***	19.1953	3.1111 (0.652)***	22.4459	3.0898 (0.638)***	21.9736	1.2796 (0.755)!	3.5951
ln(N_GE)	-0.2104 (0.066)**	0.8103	-0.2218 (0.070)**	0.80110	-0.2154 (0.068)**	0.8063	-0.0759 (0.073)	0.9270
ln(N_TS)	-1.6047 (0.523)**	0.2009	-1.8181 (0.580)**	0.1623	-1.8964 (0.571)**	0.1501	-2.1717 (0.668)**	0.1140
IMPORTS	-0.4840 (0.444)	0.6163	-0.5315 (0.450)	0.5877	-0.4881 (0.443)	0.6138	-0.3051 (0.444)	0.7371
ln(DEMANDVOL)	-0.5966 (0.042)***	0.5507	-0.6116 (0.042)***	0.5425	-0.6081 (0.042)***	0.5444	-0.5177 (0.058)***	0.5959
ln(PRODUCTPRICE)	-0.2322 (0.050)***	0.7928	-0.2680 (0.051)***	0.7649	-0.2572 (0.051)***	0.7732	-0.0631 (0.052)	0.9389
A	-3.4056 (2.393)	0.0332	-3.2243 (2.491)	0.0398	-3.1291 (2.449)	0.0438	-0.2805 (2.827)	0.7554
B	1.8223 (1.546)	6.1865	2.1866 (1.6464)	8.9045	2.2052 (1.644)	9.0722	1.8047 (1.067) !	6.0781
C	-2.2483 (1.216)!	0.1056	-2.3171 (1.232)!	0.0986	-2.2800 (1.217)!	0.1023	-0.2721 (1.352)	0.7618
D	4.0592 (0.842)***	57.9291	4.2768 (0.884)***	72.0131	-4.1767 (0.854)***	65.1528	1.8534 (0.893)*	6.3816
(0,t0]	–	–	–	–	–	–	-2.2455 (1.516)	0.1059
(t0,t1]	–	–	–	–	–	–	-2.8780 (1.528)!	0.0562
(t1,t2]	–	–	–	–	–	–	-3.7825 (1.541)*	0.0228
(t2,t3]	–	–	–	–	–	–	-5.0872 (1.516)***	0.0062
(t3,t4]	–	–	–	–	–	–	-6.1063 (1.456)***	0.0022
(t4,t5]	–	–	–	–	–	–	-6.7779 (1.522)***	0.0011
(t5,t6]	–	–	–	–	–	–	-9.6622 (1.634)***	0.0001
Intercept: $\widehat{\beta_0}$(SE)	–		-8.4685 (2.259)***		-7.6572 (2.213)***		–	
Ancillary parameter: ρ, γ (SE)	–		1.4842 (0.071)***		0.0575 (0.012)***		–	
Log-likelihood (AIC)	-1522.4 (3072.7)		-689.7 (1411.3)		-702.4 (1436.8)		-406.7 (855.4)	

Notes: PH = proportional hazard; HR = hazard ratio = exp($\hat{\beta}_1$); SE = heteroskedastic-robust standard errors; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; (0,t0]…(t5,t6] = discrete-time-interval dummies; AIC = Akaike Information Criterion.

Appendix 2E: Regression analysis on product exit – controlling for unobserved heterogeneity

Dependent variable: Product exit						
Model specified:	Cox-PH model	Weibull model	Gompertz model	Cloglog PH model		
Coefficients:	$\hat{\beta}_1$(SE)	$\hat{\beta}_1$(SE)	$\hat{\beta}_1$(SE)	$\hat{\beta}_1$(SE)[a]	$\hat{\beta}_1$(SE)[b]	$\hat{\beta}_1$(SE)[c]
COSTREDINNOV	-3.7536 (3.943)	-3.5224 (4.044)	-3.1976 (4.026)	0.8106 (4.181)	-4.1846 (15.469)	-0.2204 (9.619)
YRSINDEXP	0.0763 (0.057)	0.0861 (0.059)	0.0832 (0.058)	0.0181 (0.051)	-0.1040 (0.184)	-0.0826 (0.097)
COSTREDINNOV*YRSINDEXP	0.0589 (0.088)	0.0515 (0.091)	0.0494 (0.090)	0.0353 (0.088)	0.2026 (0.287)	0.1134 (0.152)
PRODUCTAGE	-0.0760 (0.020)***	-0.0760 (0.020)***	-0.0878 (0.019)***	0.2428 (0.872)***	0.3519 (0.169)*	0.2935 (0.111)**
ln(N_FIRMS)	2.6887 (0.684)***	2.7630 (0.772)***	2.7635 (0.766)***	0.7982 (0.872)	1.6941 (2.041)	3.1174 (1.391)*
ln(N_GE)	-0.2172 (0.065)**	-0.2301 (0.065)***	-0.2336 (0.065)**	-0.0780 (0.074)	0.5323 (0.356)	-0.0892 (0.208)
ln(N_TS)	-1.3372 (0.647)*	-1.4200 (0.761)!	-1.5120 (0.757)*	-1.7565 (0.867)*	-3.1983 (2.693)	-4.4599 (1.601)**
IMPORTS	-0.4836 (0.406)	-0.5309 (0.405)	-0.4876 (0.405)	-0.3838 (0.466)	-3.1459 (2.118)	-0.5967 (1.271)
ln(DEMANDVOL)	-0.5981 (0.039)***	-0.6139 (0.039)***	-0.6103 (0.039)***	-0.5124 (0.046)***	-0.6625 (0.155)***	-0.6968 (0.134)***
ln(PRODUCTPRICE)	-0.2361 (0.045)***	-0.2738 (0.044)***	-.26282 (.0439)***	-0.0546 (0.049)	0.2008 (0.160)	0.1354 (0.117)
A	-3.2584 (2.969)	-3.1438 (3.048)	-3.0854 (3.031)	-0.0135 (3.464)	0.6808 (10.187)	-0.3587 (9.247)
B	1.9218 (1.245)	2.2799 (1.276)!	2.2899 (1.275)!	1.9285 (1.100)!	1.7787 (4.201)	2.5370 (2.568)
C	-1.9199 (1.174)	-1.8953 (1.242)	-1.8859 (1.234)	0.3162 (1.324)	-3.7886 (6.908)	-3.8139 (4.566)
D	3.8401 (0.971)***	3.9973 (1.023)***	3.9129 (1.017)***	1.3599 (1.105)	4.2836 (3.842)	2.9925 (2.437)
(0,t0]	–	–	–	-2.1886 (1.850)	-0.3516 (6.662)	-0.3862 (3.606)
(t0,t1]	–	–	–	-2.8375 (1.858)	-3.3197 (6.372)	-1.8735 (3.572)
(t1,t2]	–	–	–	-3.8628 (1.854)*	-5.3721 (6.234)	-4.1532 (3.498)
(t2,t3]	–	–	–	-5.1196 (1.855)**	-7.7794 (6.174)	-5.9550 (3.536)!
(t3,t4]	–	–	–	-6.1796 (1.830)***	-8.9142 (6.127)	-8.2742 (3.638)*
(t4,t5]				-6.9102 (1.869)***	-8.7032 (6.427)	-7.6864 (3.688)*
(t5,t6]				-9.7878 (1.935)***	-12.8906 (6.426)*	-12.2111 (5.634)*
Intercept: $\widehat{\beta_0}$(SE)	–	-8.3810 (1.982)***	-7.5667 (1.963)***	–	–	–
Log-likelihood (AIC)	-1521.8 (3071.5)	-688.7 (1411.4)	-701.5 (1437.0)	-405.1 (854.1)	-184.6 (413.1)	-162.2 (370.3)
σ_μ^2 (Chi-square for H_0: $\mu_i = 0$)	0.0376 (1.18)	0.0534 (1.91)!	0.05018 (1.75)!	0.0747 (3.21)*	6.5750 (444.2)***	–

Notes: SE = heteroskedastic-robust standard errors; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; !$p < 0.10$; [a] Normal-distributed unobserved heterogeneity (μ_i); [b] Gamma-distributed unobserved heterogeneity; [c] Finite-mixture unobserved heterogeneity; (0,t0]…(t5,t6] = discrete-time-interval dummies; H_0 = null hypothesis; AIC = Akaike Information Criterion.

Appendix 2F: Regression analysis on price variation (using individual product prices)

Dependent variable: ln(PRODUCTPRICE)					
Model specified:	Fixed-effects model	Random-effects model	Population-averaged model	Mixed-effects linear model	
Coefficients:	$\hat{\beta}_{FE}$(SE)	$\hat{\beta}_{RE}$(SE)	$\hat{\beta}_{PA}$(SE)	$\hat{\beta}_{ME}$(SE)[a,c]	$\hat{\beta}_{ME}$(SE)[b,c]
COSTREDINNOV	-3.1883 (1.131)*	-3.3181 (0.739)***	-3.3538 (0.740)***	-3.3181 (0.740)***	-3.3181 (0.740)***
YRSINDEXP	0.0119 (0.010)	-0.0212 (0.009)*	-0.0217 (0.009)*	-0.0212 (0.009)*	-0.0212 (0.009)*
COSTREDINNOV*YRSINDEXP	0.0373 (0.019)!	0.0465 (0.017)**	0.0475 (0.017)**	0.0465 (0.015)**	0.0465 (0.015)**
PRODUCTAGE	-0.0112 (0.018)	-0.0106 (0.016)	-0.0103 (0.016)	-0.0106 (0.012)	-0.0106 (0.012)
PRODUCTAGE^2	-0.0003 (0.001)	-0.0003 (0.001)	-0.0004 (0.001)	-0.0003 (0.001)	-0.0003 (0.001)
ln(N_FIRMS)	0.2174 (0.128)	0.2718 (0.132)*	0.2705 (0.132)*	0.2718 (0.137)*	0.2718 (0.137)*
ln(N_GE)	-0.4769 (0.025)***	-0.4748 (0.025)***	-0.4745 (0.025)***	-0.4748 (0.021)***	-0.4748 (0.021)***
ln(N_TS)	-0.1342 (0.098)	-0.0065 (0.108)	0.0067 (0.112)	-0.0065 (0.154)	-0.0065 (0.154)
IMPORTS	0.9624 (0.159)***	0.9423 (0.158)***	0.9405 (0.158)***	0.9423 (0.162)***	0.9423 (0.162)***
ln(DEMANDVOL)	-0.0579 (0.006)***	-0.0581 (0.006)***	-0.0582 (0.006)***	-0.0581 (0.005)***	-0.0581 (0.005)***
A	1.4620 (0.591)*	1.2916 (0.612)*	1.2539 (0.624)*	1.2916 (0.731)*	1.2916 (0.731)*
B	1.4906 (0.202)***	0.8582 (0.226)***	0.8417 (0.225)***	0.8582 (0.239)***	0.8582 (0.239)***
C	0.4099 (0.196)*	0.1958 (0.202)	0.1915 (0.200)	0.1958 (0.327)	0.1958 (0.327)
D	0.3277 (0.196)	0.2970 (0.193)	0.2798 (0.196)	0.2970 (0.327)	0.2970 (0.327)
Intercept $\widehat{\beta_0}$(SE)	3.0685 (0.354)***	4.0042 (0.231)***	4.0220 (0.230)***	4.0042 (0.327)***	4.0042 (0.327)***
$(\widehat{\sigma_{\mu 0}}, \widehat{\sigma_{\epsilon}}; \widehat{\sigma_{\mu 1}})$	(0.249, 1.224)	(0, 1.224)	–	(4.97E-9, 1.222)	(1.06E-8, 1.222; 9.59E-9)
Covar$(\widehat{\mu_{0t}}, \widehat{\mu_{1t}})$, corr$(\hat{\beta}_0, \hat{\beta}_{1t})$	–	–	–	–	< 0.0000001, 0.9088
Variance partition coefficient	0.0398	0	–	< 0.0005	< 0.0000001
Overall R-squared	0.2140	0.2366	–	–	–
Corr$(\mathbf{X}_{it}, \widehat{\mu_{0t}})$	-0.2691	–	–	–	–
Log-likelihood (AIC)	-6013.1 (12054.2)	–	–	-6025.9 (12085.7)	-6025.9 (12081.7)

Notes: SE = heteroskedastic-robust standard errors; ***p < 0.001; **p < 0.01; *p < 0.05; !p < 0.10; [a] Without a random-slope for COSTREDINNOV; [b] With a random-slope for COSTREDINNOV; [c] Log-likelihood ratio test for a comparison with the equivalent single-level model was not statistically significant. LR test for a comparison of (a) and (b) indicates no statistically-significant evidence that the effect of COSTREDINNOV is not fixed: it does not vary across different time periods; AIC = Akaike Information Criterion.

Appendix 3A: Database searches

CRD databases	**Hits**
Search date: 16th April 2013	
Results: 1066	
1. MeSH DESCRIPTOR economics EXPLODE ALL TREES	13233
2. ((drug* or route* or frequenc* or administr* or deliver* or form* or dos*))	30648
3. (cost* or cost analysis*)	19664
4. #2 AND #3	12616
5. ((biologic* or biopharmaceu* or protein* or macromolecul*))	3197
6. ((fab* or mab* or immunoglobul* or antibod* or fragment* or monoclonal*) or protein* or macromolecul*))	3778
7. #4 AND #5 AND #6	628
8. #1 AND #7	459
9. (cost* adj2 (administr* or form* or delivery* or route*))	644
10. #8 OR #9	**1066**

EMBASE Classic + EMBASE (1947 to 2013 Week 15)	**Hits**
Search date: 15th April 2013	
Results: 692	
1. (cost* adj5 (drug* or administ* or route* or delivery* or form* or dos*)).ti,ab.	26575
2. (cost* or administ* or form* or delivery* or route*).mp.	5886547
3. (drug* or route* or frequenc* or administ* or delivery* or form* or dos*).mp.	10157799
4. (biologic* or biopharmaceu* or protein* or macromolecul*).mp.	4555850
5. (fab* or mab* or immunoglobul* or antibod* or fragment* or monoclonal* or protein* or macromolecul*).mp.	5066634
6. #2 and #3 and #4 and #5	1057218
7. exp health economics/	581523
8. #6 and #7	11870
9. #1 and #8	**692**

MEDLINE (OvidSP)	**Hits**
Search date: 15th April 2013	
Results: 2448	
1 economics/	26558
2 exp Economics, Pharmaceutical/	2416
3 Economics, Medical/	8493
4 Economics, Hospital/	9715
5 Economics, Nursing/	3870
6 (cost* adj5 (drug* or administ* or route* or delivery* or form* or dos*)).ti,ab.	18370
7 (cost* or administ* or form* or delivery* or route*).mp.	3931359
8 (drug* or route* or frequenc* or administ* or delivery* or form* or dos*).mp.	5871197
9 (biologic* or biopharmaceu* or protein* or macromolecul*).mp.	4042919
10 (fab* or mab* or immunoglobul* or antibod* or fragment* or monoclonal* or protein* or macromolecul*).mp.	4129502
11 #7 and #8 and #9 and #10	849792
12 #6 and #11	1268
13 #9 or #10	4879209
14 #7 or #8	6130325
15 #6 and #13 and #14	**2448**

Econlit (EBSCO Host)	**Hits**
Search date: 15th April 2013	
Results: 138	
S1 (cost* or administ* or form* or delivery* or route*)	259618
S2 (drug* or route* or frequenc* or administ* or delivery* or form* or dos*)	166151
S3 (biologic* or biopharmaceu* or protein* or macromolecul*)	2808
S4 ((fab* or mab* or immunoglobul* or antibod* or fragment* or monoclonal*) or protein* or macromolecul*)	11777
S5 (S1 or S2)	272272
S6 (S3 and S4)	443
S7 (S5 and S6)	**138**

Appendix 3B: Summary of studies reviewed[i]

Study	Description	Components of administration costs	Type of cost	Cost of administration
Bravo et al.[75]	CEA/CUA of two TNF antagonists (s.c. etanercept and i.v. infliximab) and palliative care in people with PsA who have failed on DMARDs.	Staff nurse time (£40.50 per patient-related hour); first attendance (£131) and follow-up (£94) attendance in outpatient rheumatology; and day-case rheumatology (£614).	Both, i.e., PAc plus Pc	Over an initial 3-month period, administration cost of etanercept = £293; £920 for infliximab infusion.[ii]
Cassidy et al.[69]	CEA/CUA comparing oral capecitabine (Xeloda®) with bolus i.v. 5-FU/LV in people with Duke's C (stage III) colon cancer and have undergone resection of the primary tumour.	Physician consultation (unit cost of £68) and costs of infusion per visit (intravenous administration = £201) over a 24 week period.	Both	£499 = administration costs for oral capecitabine; £6136 = administration costs for 5-FU/LV.
Chen et al.[60]	CEA/CUA of s.c. etanercept, i.v. infliximab or s.c. adalimumab against DMARDs in adults with RA who are not well controlled by conventional DMARDs.	Infusion cost for infliximab. Administration costs for etanercept and adalimumab were not reported.	PAc	£148 = costs per infliximab infusion.
Cummins et al.[62]	CEA/CUA of treatment sequencing strategy of TNF antagonist followed non-biologic DMARDs versus palliative care without TNF inhibitors followed by non-biologic DMARDs in people with active and progressive PsA.	Intravenous infusion cost for infliximab. Subcutaneous formulations adalimumab and etanercept require initial consultant outpatient visit followed by two nursing visits to educate patients on self-administration.	Both	£128 = infliximab costs per infusion; £408 = self-administration costs for adalimumab and etanercept for first cycle.
Dretzke et al.[99]	CEA/CUA of induction and maintenance therapy with s.c. adalimumab and i.v. infliximab (relative to standard care) in the management of people with moderate to severe active CD.	Cost of having an infliximab infusion in a hospital covering the cost of a health professional during the 2 hours of infusion and for the period of time afterwards.	PAc	£285 = infliximab infusion costs per treatment. No administration cost was assigned to adalimumab.
Drummond et al.[58]	Cost consequence analysis comparing two anticoagulant therapies, standard (unfractionated) heparin and enoxaparin for the prophylaxis of postoperative deep	For both i.v. standard heparin and s.c. enoxaparin, this includes the costs of consumables (needles and syringes) and 10 minutes of nursing time per injection.	PAc	£104 per standard heparin injection; £26 per enoxaparin injection.

	vein thrombosis and pulmonary embolism in people undergoing elective hip surgery.			
Holmes et al.[70]	CEA/CUA of docetaxel as second-line treatment versus BSC in people with NSCLC who have received prior treatment with a platinum (cisplatin or carboplatin) containing chemotherapy but had not been previously treated with taxanes.	30 minutes of nurse time (costed at £45 per hour) for a one-hour i.v. infusion of docetaxel.	PAc	£23 = cost per docetaxel infusion.
Hoyle et al.[71]	CEA/CUA comparing i.v. temsirolimus to s.c. interferon alfa-2a for first line treatment of people with stage IV or recurrent renal cell carcinoma who have not received any prior systemic therapy.	District nurse visit and NHS HRG code: "Deliver subsequent elements of a chemotherapy cycle."	Both	£26 = administration costs of interferon alfa (cost of community or district nurse visit). £203 = administration cost of temsirolimus per i.v. in the hospital.
Iveson et al.[100]	CEA/CUA of replacing i.v. 5-FU either as a single agent (Lokich regimen [B2]) or in combination with folinic acid (de Gramont regimen [B1] or AIO regimen [B3]) with i.v. irinotecan as second-line therapy for metastatic colorectal cancer.	Staff time; hospitalization for chemotherapy administration; insertion of a tunnelled central line catheter; use of disposable pumps for each treatment cycle.	Both	£495 = per patient administration costs for 6 cycles of irinotecan. £2154 = per patient administration costs of 6.8 cycles of B1 regimen (= 1 catheter insertion and 6.8 disposable pumps). £2748 = per patient administration costs of B2 regimen (= 13.3 weeks of hospital day attendance, 1 catheter insertion and 13.3 disposable pumps). £3460 = average of per patient administration costs for B3 regimen (requiring inpatient care or hospital day attendance)[iii].
Jobanputra et al.[61]	CEA/CUA of introducing TNF antagonists into the treatment strategy for adults with RA who are not well	Infliximab i.v. infusion.	PAc	£175 = cost per infliximab infusion.

	controlled by conventional DMARDs.			
Kielhorn et al.[66]	CEA/CUA of i.v. rituximab in people with RA who have not responded adequately to at least two non-biologic DMARDs and one or more TNF antagonists.	Cost of i.v. infusions or injections. For infliximab, this was 3 hours of infusion time including cost of post-infusion observation. For rituximab, 5 hours of infusion was assumed plus costs of premedication.	Both	Average annual administration cost per patient: £369 for rituximab; £788.50 for infliximab; £54 for adalimumab. Administration costs for the non-biologic DMARDs assumed to be zero.
Knight et al.[67]	CEA/CUA of adding i.v. rituximab to CHOP versus CHOP alone for adults with DLBCL.	Pharmacy cost of dispensing and doctor/nurse visits.	Both	£140 = administration costs of CHOP (= pharmacy dispensing cost per cycle [£36] + doctor/ nurse costs per cycle [£104]). £125 = cost of administering rituximab (= pharmacy dispensing cost per cycle [£12] + doctor/ nurse costs per cycle [£113]).
Le et al.[59]	CMA of oral and intravenous chemotherapies for first line treatment of people with advanced NSCLC. These include different doses of oral and i.v. VNB and i.v. chemotherapies containing gemcitabine, paclitaxel and docetaxel.	Day-time hospitalization (£422) and medical oncology outpatient visit (£143). For self-administration of an oral agent at home, a cost for pre-therapy counselling with hospital nurse (£18) is incurred plus the cost of a local GP (£24) to carry out blood tests.	Both	*Annual administration costs*: **60mg oral VNB** = £845 (= outpatient visit [£505] + home care [£340]); **80mg oral VNB** = £1054 (= outpatient visit [£610] + home care [£444]). **25mg i.v. VNB** = £1115 (= inpatient stay [£697] + outpatient visit [£418]); **30mg i.v. VNB** = £1189 (= inpatient stay [£734] + outpatient visit [£455]). **1000mg i.v. gemcitabine** = £1463 (= inpatient stay [£871] + outpatient visit [£592]); **1250mg i.v. gemcitabine** = £1618 (= inpatient stay [£948] + outpatient visit [£670]). **100mg i.v. docetaxel** = £3922

				(= inpatient stay [£2100] + outpatient visit [£1822]). **175mg i.v. paclitaxel** = £3438 (= inpatient stay [£1858] + outpatient visits [£1580]); **200mg i.v. paclitaxel** = £4033 (= inpatient stay [£2156] + outpatient visit [£1876]).
Lewis et al.[72]	CEA/CUA comparing oral erlotinib with i.v. docetaxel in the second line treatment of people with previously treated stage IIIb-IV NSCLC.	NHS HRG code: "Deliver simple parenteral chemotherapy at first attendance" = £209	Both	£209 = docetaxel administration per visit. Zero administration costs assigned to oral erlotinib.
Loveman et al.[101]	CEA/CUA of topotecan compared to existing regimens in second-line chemotherapy for small cell lung cancer.	Pharmacy cost for dispensing and outpatient attendance to receive oral topotecan; outpatient attendance to receive chemotherapy, pharmacy preparation cost (i.v. topotecan)[iv].	Both	£200 = administration costs for oral topotecan per cycle; £1059 = administration costs per cycle for i.v. topotecan
Morris et al.[91]	CEA/CUA comparing recombinant activated factor VII (rFVIIa, eptacog alfa, NovoSeven®) given as i.v. bolus injection with placebo for the control of bleeding in people with severe blunt trauma.	None reported. Each NovoSeven® pack contains 1 vial of the drug, 1 vial of sterile water, 1 sterile vial adapter, sterile syringe and infusion set, and swabs.	Not specified	It appears an assumption of zero administration costs was made.
Robinson et al.[73]	CEA/CUA comparing i.v. GC with MVAC combination therapy given by i.v. in the treatment of people with locally advanced or metastatic bladder cancer.	Cost per day of inpatient chemotherapy delivery = £355. Cost per outpatient attendance for chemotherapy delivery = £292.41.	Both	£3639 = per patient administration costs of GC inpatient; £5429 per patient if administered in outpatient settings. £3157 = per patient administration of MVAC inpatient; £4522 if administered in outpatient settings.
Starling et al. [102]	CEA/CUA comparing i.v. cetuximab plus irinotecan against BSC for the treatment of people with metastatic colorectal cancer who failed on previous chemotherapy.	Not reported.	Not specified	£304 per day = administration costs for cetuximab plus irinotecan.

Sweetenham et al.[68]	CEA/CUA comparing i.v. CHOP with a purine analogue, i.v. fludarabine, and an i.v. anti-CD20 monoclonal antibody, rituximab, in the treatment of people with relapsed indolent B-cell non-Hodgkin's lymphoma.	Tests, inpatient stays (£382 per day) and outpatient visits (£98 per day) for drug administration.	Both	£1271 = administration costs for rituximab per patient per treatment course; £1725 = administration costs for CHOP; £3939 = administration costs of fludarabine.
Ward et al.[76]	CEA comparing gemcitabine with two 5-FU regimens (de Gramont regimen and PVI 5-FU) in first- and second-line treatment of people with pancreatic cancer.	Outpatient visits for gemcitabine infusion. Administration of PVI 5-FU involves insertion and removal of central line (as day case), and the cost of checking and flushing a central line and ambulatory pump.	Both	£153.50 = per month cost of medical oncology outpatient visit for gemcitabine administration. £1181 = total administration costs of PVI 5-FU per month[v].
Woolacott et al.[74]	CEA/CUA of three treatment options (s.c. etanercept, i.v. infliximab and palliative care) in people with PsA who have failed DMARDs.	As reported in Bravo et al.[23]	Both	As reported in Bravo et al.[23].
Colquitt et al.[103]	CEA/CUA of CSII with the aid of an insulin pump versus MDI in people with type I diabetes.	Acquisition and maintenance costs of insulin pump, insulin cartridges/ reservoirs, infusion sets and batteries. Three types of insulin pumps considered: Disetronic H-Tron, Disetronic D-Tron and MiniMed 508.	PAc	£5196 = net administration cost of Disetronic H-Tron; £4785 = net administration cost of Disetronic D-Tron; and £4826 = net administration cost of MiniMed 508.
Cummins et al.[104]	CEA/CUA of CSII with the aid of an insulin pump versus MDI in people with type I diabetes.	Acquisition and maintenance costs of insulin pumps and cost of consumables consisting of infusion sets, insulin reservoir, needles, lancets, test strips and glucometer.	PAc	£2319 = annual cost of CSII £513.27 = annual cost of MDI.[vi]
Cummins et al.[105]	CEA/CUA of s.c. golimumab and other TNF inhibitors given subcutaneously in people with active PsA.	Outpatient visits and staff nurse time. Drug monitoring costs estimated separately.	Both	Administration cost = £341 (1 outpatient visit + 4 hours of nurse time) over an initial 12 week period; £94 (1 outpatient visit) over the next 12 weeks.

Hoyle et al.[77]	CEA/CUA comparing dasatinib and nilotinib versus high-dose imatinib in people with chronic phase CML, who are resistant to normal dose imatinib; and dasatinib and nilotinib with interferon alfa in people with chronic phase CML, who are intolerable to imatinib.	District nurse visit for s.c. interferon alfa administration in those not able to self-administer. Oral treatments (dasatinib, nilotinib and high-dose imatinib) incur no administration costs.	PAc	£28 = administration cost of interferon alfa. Average cost per 2-month cycle = £421.
Lloyd et al.[78]	CMA comparing s.c. highly purified hMG and s.c. rFSH in women aged 18 to 36 years with infertility for more than one year and who are undergoing *in vitro* fertilization or intracytoplasmic sperm injection.	Staff costs of administering injections plus cost of clinic visit.	Both	£53.53 = cost of administering injections (£25.33) + cost of clinical visit (£28.20).
Martin et al.[79]	CEA/CUA of providing antiviral treatment (s.c. peginterferon plus oral ribavirin) for IDUs at risk of hepatitis C virus transmission compared with treating ex/non-IDUs or no treatment.	Staff costs and tests related to evaluations, further investigation visits, treatment appointments, assessments, sustained-viral-response surveillance and IDU psychiatric visits.	Both	£2821.14 = total delivery costs per treatment over a 48 week period.
Pinot et al.[86]	CEA of screening UK-born and Bangladesh-born first-time pregnant women for varicella infection and then vaccinating those who are susceptible.	Nurse consultation (two nurse consultations needed for the two adult doses given s.c.).	PAc	£13 = vaccine administration costs.
Palmer et al.[106]	CEA/CUA comparing s.c. IDet-based basal/bolus therapy with NPH-based basal/bolus insulin in people with type I diabetes.	Costs of needles and devices. Unit cost of NovoPen 3® device = £28.60 and needles (NovoFine®) = £14.40 per pack of 100.	PAc	£62 = annual cost of needles and devices for both IDet- and NPH-based insulin therapy.
Rogers et al.[107]	CEA/CUA comparing dasatinib and nilotinib in people with CML in chronic phase.	As reported in Hoyle et al.[25]	PAc	As reported in Hoyle et al.[25]
Wolowacz et al.[80]	CEA/CUA of oral dabigatran etexilate compared with s.c. low-molecular-weight heparin (LMWH, also enoxaparin) for the	Inpatient and outpatient visits for administration of LMWH. Inpatient visits comprise of nurse time for injection of	Pc (outpatient); Both (inpatient)	Zero costs for oral dabigatran; £0.85 = cost of inpatient visit (both TKR and THR); £26 =

	prevention of venous thromboembolism in people undergoing TKR and THR surgery.	LMWH[vii].		cost of outpatient visit (for only THR).
Chilcott et al.[81]	CEA/CUA comparing three AADP regimens for immunization against haemolytic disease of the newborn in primagravidae and all pregnant women who are rhesus D negative.	Cost of midwife administering i.m. anti-D IgG during a normal antenatal visit.	PAc	£12 = cost of administering two doses of anti-D IgG per pregnancy.
Fenn et al.[82]	CEA/CUA of vaccination programmes for hepatitis B in infants, children aged 6 years and above, and adolescents aged 11 years and above.	Cost of GP consultation except for the primary infant vaccination schedule. Doses given i.m.	PAc	£20 = vaccine administration costs per dose (applies to both initial and booster doses).
Jit et al.[84]	CEA/CUA of vaccinating women in their second or third trimester of pregnancy against seasonal influenza in England and Wales.	Item of service payment to GPs. Doses given i.m.	PAc	£7.78 = vaccine administration cost per person vaccinated.
Mangtani et al.[108]	CEA/CUA of adding universal hepatitis B vaccination for infants and preadolescents to a policy of selective vaccination of adults.	Staff costs (nursing time) for i.m. dose administration in all programs. Cost of needles and syringes applies to only selective adult and universal pre-adolescent vaccination. Cost of visits for second or third dose applies to only selective adult vaccination.	Both	£4.3 = administration cost per dose for universal infant vaccination. £8.70 = administration cost per dose for universal preadolescent vaccination. £51 = administration cost per dose for selective adult vaccination.
McIntosh et al.[85]	CEA/CUA of 7-valent PCV in infants and young children ≤ 2 years of age.	Not reported. Doses given i.m.	Not specified	£13 = vaccine administration cost per child or infant vaccinated.
Nuijten et al.[89]	CEA/CUA of i.m. palivizumab, a preventive therapy for severe respiratory syncytial virus infection versus no prophylaxis in children at high risk of hospitalization, i.e. preterm infants of ≤ 35 weeks gestation, children with bronchopulmonary dysplasia and children	Visit to a hospital outpatient facility, which includes a specialist fee.	Both	£161 = cost of outpatient visit.

	with congenital heart disease.			
Siddiqui et al.[83]	CEA/CUA of universal infant or adolescent vaccination of hepatitis B as well as selective vaccination of ethnic infant populations.	Not reported. Doses given i.m.	Not specified	£5.40 = administration costs per dose (for adolescent vaccination programme). Zero costs for the infant vaccination programme.
Trotter et al.[85]	CEA/CUA of MCC vaccination programme in infants and children < 18 years of age in England and Wales.	Payment for vaccine delivery (by GP or in a school). Doses given i.m.	PAc	£8.80 = vaccine delivery payment per dose if administered by a GP (£6.06 per dose if more than one dose is required, in which case the fee charged is for all doses with the exception of the last dose in the series). £1.4 if administered in schools.
Vick et al.[88]	CEA/CUA of three regimens of routine AADP in populations at risk of having babies with Rh haemolytic disease.	Midwife's time and consumables used. AADP given i.m.	PAc	£4.40 = administration costs per injection.
Jit et al.[90]	CEA/CUA of rotavirus vaccination in five European countries (Belgium, England and Wales, Finland, France and the Netherlands) in infants and young children.	10 minutes of practice nurse time (assuming administration in GP clinics). Doses given orally.	PAc	£6.10 = vaccine administration cost per dose.
Lowin et al.[109]	CEA/CUA of levodopa/carbidopa intestinal gel versus standard care in people with advanced Parkinson's disease at Hoehn and Yahr stages 3, 4 or 5 experiencing more than 50% of waking time in the OFF state[viii].	Temporary insertion of NG tube requiring one day of inpatient stay plus permanent insertion of PEG requiring one day of inpatient stay.	PAc	£987 = costs of NG tube insertion (£204) + cost of PEG tube insertion (£783).
Walleser et al.[110]	CEA/CUA of first-line maintenance erlotinib versus best supportive care in people with locally advanced and metastatic NSCLC.	Not reported.	Not specified	Zero administration costs as erlotinib is given orally.

[i] Abbreviations used: CEA = cost-effectiveness analysis; CUA = cost-utility analysis; TNF = tumour-necrosis-factor; s.c. = subcutaneous; i.v. = intravenous; PsA = psoriatic arthritis; DMARDs = disease-modifying antirheumatic drugs; PAc = physical administration costs; Pc = proximal costs; 5-FU/LV = 5-fluorouracil with leucovorin; RA = rheumatoid arthritis; CD = Crohn's disease; BSC = best supportive care; NSCLC = non-small-cell lung cancer; NHS = National Health Service; HRG = healthcare resource group; 5-FU = 5-fluorouracil; AIO = Arbeitsgemeinschaft Internistische Onkologie; CHOP = a combination therapy of oral cyclophosphamide, intravenous doxorubicin, intravenous vincristine and oral prednisone; DLBCL = diffuse large B-cell lymphoma; CMA = cost minimization analysis; VNB = vinorelbine; GP = general practitioner; GC = gemcitabine plus cisplatin; MVAC = methotrexate, vinblastine, doxorubicin and cisplatin; PVI 5-FU = protracted venous infusion of 5-fluorouracil; CSII = continuous subcutaneous insulin infusion; MDI = multiple daily injections; CML = chronic myeloid leukemia; hMG = human menopausal gonadotropin; rFSH = recombinant follicule stimulating hormone; IDUs = injecting drug users; IDet = insulin detemir; NPH = neutral protamine Hagedorn; LMWH = low-molecular-weight heparin; TKR = total knee replacement; THR = total hip replacement; AADP = antenatal anti-D prophylaxis; i.m. = intramuscular; IgG = immunoglobulin G; PCV = pneumococcal conjugate vaccine; MCC = meningococcal serogroup C conjugate; NG = nasogastric; PEG = percutaneous endoscopic gastronomy.

[ii] Cost of a single infliximab infusion is equal to half rheumatology day case (3 day case attendances needed over the initial 3 months). Within the initial 3 month period, administration cost of etanercept includes educational visit for etanercept self-injection (equal to one outpatient rheumatology attendance) and four visits to a staff nurse to check on treatment progress. Over 12 months, infliximab infusion costs = £1994, corresponding to 6.5 of half rheumatology day cases.

[iii] It is assumed that 50% of patients will receive B3 regimen inpatient with the other 50% receiving the treatment as one day hospital attendance. Administration costs of B3 regimen requiring inpatient stay = £4245 (for 12.9 cycles of inpatient stay and 1 insertion of a catheter). Administration costs of B3 regimen requiring day hospital attendance = £2676 (for 12.9 cycles of day hospital attendance and 1 insertion of a catheter).

[iv] The study estimates the cost of chemotherapy administration separately from on-treatment monitoring costs. Oral and i.v. topotecan incur the same on-treatment monitoring costs (£95) covering full blood count, liver function tests, renal function tests, chest radiograph and computed tomography scan every two cycles.

[v] The monthly administration cost of PVI 5-FU consists of £235 in infusion costs (covering insertion and removal of a central line/pump and flushing of the central line/pump), £711 in costs of hospitalization for drug administration, operational costs of £209 and £26.10 for concomitant medications.

[vi] Annual cost of CSII includes the cost of infusion sets (£1,058.87); insulin reservoirs (£325.82); insulin cartridge (£312.21) and annualized capital cost of insulin pump of £620 (assuming a 4-year pump life). Annual cost of MDI includes the cost of needles (£31.83), insulin cartridge (£466.44) and annualized capital cost of pen devices of £15 (assuming a three-year lifespan). The costs of lancets, test strips and glucometer are the same for both CSII and MDI.

[vii] Cost of outpatient visits incurred by 13% of patients unable to self-administer refer to one district nurse visit per day. For the 87% able to self-administer, a one-time cost (= 30 minutes of nurse time [£11.92]) for training in self-administration is incurred during hospital stay. This gives an average administration cost for LMWH per injection of £13.75.

[viii] OFF state refers to the period where effects of standard oral treatments for Parkinson's disease wear off and symptoms return.

Appendix 5A: Demographic characteristics of the UK sample

Characteristics (of 442 respondents)	N (% of sample)
Gender	
Male*	181 (40.95)
Female	261 (59.05)
Respondents' age	
17 – 34 years	151 (34.16)
35 – 49 years	155 (35.07)
≥ 50 years	136 (30.77)
Employment status	
Employed	307 (69.46)
Unemployed*	135 (30.54)
Household-income category	
< £15,000 (per year)	109 (24.66)
£15,000 – £29,999 (per year)	134 (30.32)
£30,000 – £49,999 (per year)	118 (26.70)
£50,000 – £75,000 (per year)	52 (11.76)
> £75,000 (per year)*	29 (6.56)
Highest education achieved	
GCSEs O & A levels	189 (42.76)
Higher education*	212 (47.96)
Vocational training	41 (9.28)
Prior illness (in the past year)	
Yes	178 (40.27)
No*	264 (59.73)

Notes: N = number of respondents; * indicates the reference category for the effects coding used (see also Table 5.3 on page 126).

Appendix 5B: Demographic characteristics of the US sample

Characteristics (ΣN)	N (% of sample)
Gender (209)	
Male*	65 (31.10)
Female	144 (68.90)
Respondents' age (208)	
Under 50 years	105 (50.48)
≥ 50 years	103 (49.52)
Type of healthcare institution (209)	
Solo medical practice	14 (6.70)
Government-run	15 (7.18)
Private-for-profit	67 (32.06)
Non-for-profit	89 (42.58)
Other*	24 (11.48)
Patient case-mix (209)	
Inpatients	78 (37.32)
Outpatients	80 (38.28)
Accidents & emergency	22 (10.53)
Other*	29 (13.88)
US census region (206)	
East North Central	31 (15.05)
East South Central	10 (4.85)
Middle Atlantic	27 (13.11)
South Atlantic*	38 (18.45)
Mountain	16 (7.77)
New England	22 (10.68)
Pacific	24 (11.65)
West North Central	20 (9.71)
West South Central	18 (8.74)

Notes: N = number of respondents. * indicates the reference category for the effects-coding used (see also Table 5.4 on page 127-128). This dataset offered a complete sequence of 12 choices for each of the 210 survey respondents. However, for three respondents, data was missing on some of their characteristics whilst for one respondent there was no information on *all* characteristics.